Measurement of the differential $t\bar{t}$ cross section in the boosted all-hadronic channel for the CMS experiment at the LHC at $\sqrt{s} = 8$TeV

by

David Kenneth Fehling Jr.

A dissertation submitted to The Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy.

Baltimore, Maryland

January, 2017

Abstract

The first measurement of the differential $t\bar{t}$ cross section at using 19.7 fb^{-1}of all-jet events with boosted high-p_T top quark jets collected by the CMS detector at $\sqrt{s} = 8$ TeV is presented. A boosted top jet is one in which the subjets have merged. A data-driven method is used to calculate the standard model background dominated by QCD multijet production. The $t\bar{t}$ cross section is extracted in bins of measured leading top quark p_T using a binned likelihood fit of the invariant mass distribution of top jet candidates. The $t\bar{t}$ cross section as a function of measured leading top quark p_T is then unfolded to parton-level in order to facilitate comparison with theory. This is the first measurement of the differential $t\bar{t}$ cross secton in the boosted regime in the all-jet channel at $\sqrt{s} = 8$ TeV.

Primary Reader: Petar Maksimovic

Secondary Reader: Barry Blumenfeld

Acknowledgments

This work would not have been possible without the countless individuals who have been there for me and supported me through this process.

I would like to begin by thanking my advisor, Petar. Petar, you're a great scientist, but a better person. You've gone well above and beyond your responsibilities as an advisor and mentor and I'll forever appreciate it. "Better is the enemy of good enough" will be with me forever.

There have been many JHU administrators and faculty that have had the pleasure of knowing. Morris, I'll miss your openness and your laughter. To Bruce, Barry, and Andrei, the HEPex group may be small, but you were always accessible and working hard to make it succeed. Pam, you've made this department feel like home to me. Kelley, you've helped me through this hard time at the end while also constantly improving the Department: thank you. I hope you get back to San Francisco at some point. Zlatko Tesanovic, wherever you may be, you went out of our way to reassure me when I was starting out and in over my head. That meant the world to me.

During my time at Johns Hopkins, I have had the privilege of meeting and making

friends with many students. In general, you've all been great. To the original crew: Justin, Nhan, Keith and Bridget (Sarah) you were the best and really helped me grow a lot as a person. You also gave me a great nickname! Our time together was special and I'll always remember it fondly.

Nik, my kindred spirit, I hope you're well. Ian, you just might be the coolest person I know. I'll miss our talks as well as watching your gameplay. Marc, you were the worst, but in the best of ways. Your tough exterior hides a heart of gold. Nick, Nikki, Keith, Lisa, Matt and Helene, you all were there for me in the most trying of times and your friendship and understanding have been invaluable to me. You've kept me sane and focused and helped me keep going. Derek, your unique viewpoint has been refreshing. To everyone else, it's been great getting to know you all.

Lastly, but most certainly not least, I need to thank my family. Rather than keeping me grounded, you all always pushed me and had faith in me when I did not myself. To the Winters family, Gene, Stephanie, Allison, Tim, Ryan. You've all always been so nice to me and treated me so well. Your compassion for me and other people has always meant a lot to me.

Danielle, you've been a testament to perseverance. I love you and know that you'll never be kept down. Linda, you're such a caring and independent woman. I respect you so much. Michelle, you ARE the coolest person I know. You're who I want to be when I grow up. To all my nieces and nephews, Uncle David loves you all!

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Physics is the branch of science concerned with the nature and properties of matter and energy. In order to even have a chance of understanding and explaining such a massive field, Physicists usually break their study down according to a fundamental scale such as size or characteristic interaction distance. Astronomy covers the largest distances while Condensed Matter and Particle Physics study much smaller, generally sub-atomic, scales. An active area of Particle Physics involves accelerating particles (giving them more energy) and colliding them to study the properties of the collided or created particles. By knowing the intial collision conditions and understanding the results from the collision, information about the fundamental nature of matter can be gleaned. The method by which these collisions and measurements are made will be discussed in Chapter 2. This chapter will discuss the current understanding of the nature of matter as described by the Standard Model of Particle Physics.

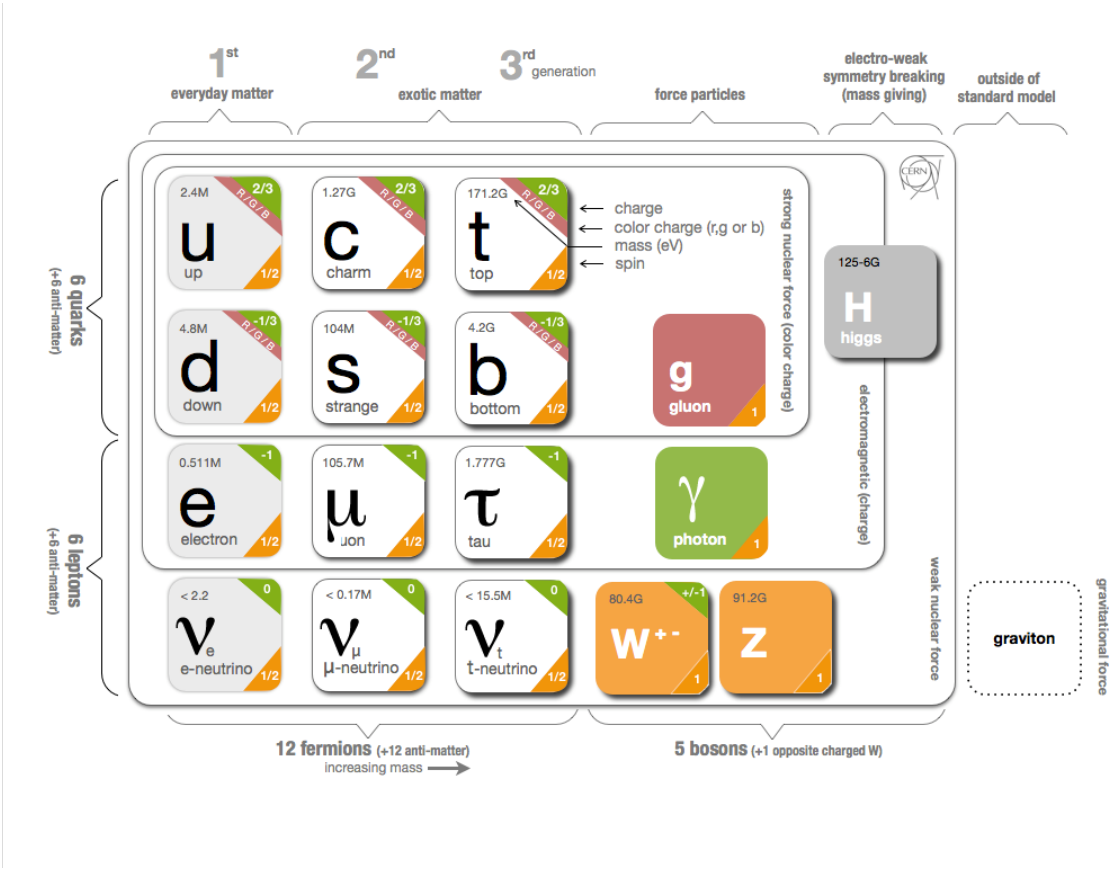

Figure 1.1: Graphical summary of the Standard Model of Particle Physics. Image taken from Ref. [1].

1.1 The Standard Model

In the 1960s and 1970s, the Standard Model of Particle Physics (SM) arose out of attempts by theorists (Glashow, Weinberg, Salam[1]) to combine and describe the interactions of matter. This theory predicted with great accuracy not only the existence of the $W^{\pm}$ and Z^0 bosons, but also their properties. Since then, the Standard Model has been repeatedly examined and experimentally tested. These experiments have (with a few exceptions[2]) all supported the SM, leading it to be one of the most well-tested models in Physics.

The Standard Model describes the components of visible matter and their interactions, save gravity. Matter is made of particles with half-integer spin known as *fermions* and mediated by particles with integer spin called *bosons*.

A graphical summary of the SM is given in Fig. 1.1. There are many good reviews of particle physics and the Standard Model, but Ref. [14] serves as a fairly accessible introduction.

1.1.1 Fermions

Fermions can be further divided into *leptons* and *quarks*. Fundamentally, there are 12 leptons and 36 quarks, and each set is usually grouped into 3 generations, ordered by increasing mass.

[1] And many, many others.

[2] Such as the anomalous magenetic moment of the muon (a_μ) or neutrino masses.

1.1.1.1 Leptons

Many people are already familiar with the first generation lepton, the *electron*. The electron has a mass of 0.511 MeV/c^2 or conventionally, 0.511 MeV[3,4,5]. The electron has electric charge of -1[6] and spin of 1/2. The *muon* (μ) and *tau* (τ) have the same properties of as the electron, but with much larger masses (more than 200 and 3400 times greater, respectively).

Each lepton has an associated *neutrino* (ν), a nearly massless particle with 0 electric charge and spin of 1/2. The neutrinos are among the lightest particles ever observed and generally don't react with matter, so an accurate measurement of their masses is quite difficult. The SM makes no prediction for the neutrino masses, so an external correction must be added to account for them. The strongest evidence for massive neutrino is *neutrino oscillations* and the 2015 Nobel Prize in Physics was awarded for this discovery.

Electron neutrinos are created in the sun by the nuclear reaction:

$$p + p \rightarrow^2_2 He + \gamma \rightarrow^2_1 Di + e^+ + \nu_e \tag{1.1}$$

[3]In particle physics it is convention to set c (the speed of light) to 1. Then mass and energy are treated identically, as shown by E = mc^2, with c = 1. Plank's constant, h, is also set to 1, equating energy and momentum.

[4]One eV (or electronvolt) is the energy gained by an electron passing through a potential of 1 Volt.

[5]The unit prefixes follow the standard SI convention, the ones relevant to this analysis are:

k	1,000 (10^3)	m	.001 (10^{-3})
M	1,000,000 (10^6)	μ	.000001 (10^{-6})
G	1,000,000,000 (10^9)	n	.000000001 (10^{-9})
T	1,000,000,000,000 (10^{12})	p	.000000000001 (10^{-12})

[6]Electric charge is generally quoted as a number between -2 and +2. In all cases, this is to be multipled by the fundamental charge of the electron (1.602×10^{-19} Coulombs).

where p is a proton, He is Helium, Di is Deuterium (an isotope of Hydrogen with neutrons instead of 1), e^+ is a positron, and ν_e is an electron neutrino. However, the amount of electron neutrinos detected on the Earth was only 1/3 of the SM prediction. The total neutrino flux observed was equal to the expected electron neutrino flux. The neutrinos must have mass in order to allow the electron neutrino to osillate to the other ones and account for the observed behavior.

In the SM, each of the six leptons described has an antiparticle and hence the total number of leptons is 12. An antiparticle has the same mass, same spin as its partner particle, but has the opposite charge. In the nuclear equation given previously, one of the final particles is an antielectron (e^+, commonly called a positron)[7]. As the neutrinos have no electric charge, it is possible that they are their own antiparticles; this is an open question in Physics.

1.1.1.2 Quarks

The remaining fundamental fermions in the Standard Model are quarks and antiquarks. As with the leptons, the quarks and their antiquark partners have the same mass and spin of 1/2, but opposite charge. The first generation (and lightest) quarks are the *up* (u) and *down* (d) quarks. The up quark has a mass of approximately[8] 2.3 MeV and electric charge of 2/3 while the down quark has a mass of 4.8 MeV and an

[7]Antiparticles are distinguished from their partner particle by either charge conjugation as in this case where e^+ is the antiparticle of e^- or the addition of a bar ($\bar{q}$ is the antiparticle of q).

[8]Due to many complications including quark confinement, the masses of the light quarks are not precisely known

electric charge of -1/3. Quark masses are extremely difficult to measure, and as a result, the values quoted for very approximate[9]. The remaining quarks are generally described as *up-type* or *down-type*[10] based on whether they have a charge of 2/3 or -1/3, respectively. The second generation down-type quark has a mass of 95 MeV and is called the *strange* (s) quark and the second generation up-type quark, the *charm* (c) quark, has a mass of 1.28 GeV. The third generation down-type quark is the *bottom* (b) quark and has a mass of 4.18 GeV. The final quark, the up-type *top* (t) quark, was discovered in 1995 and has a mass of 173 GeV.

As previously mentioned, particles and antiparticles differ in charge—for each of the two types of electric charge (+ and -) there exists a fundamental particle. *Color* is another type of charge affected by the strong force. This will be discussed in Sec. 1.1.2.3, but with regard to the number of quarks in the SM, it is enough to know that there are three types of color charge (*red*, *green*, and *blue*). Each of the six quarks (udscbt) and their antiparticle partners ($\bar{u}\bar{d}\bar{s}\bar{c}\bar{b}\bar{t}$) discussed can have one of the types of strong charge[11], for a total of 36 fundamental quarks.

Unlike leptons, quarks are not observed directly in nature, a phenomenon called *quark confinement*. They are always found in bound states consisting of two quarks and one antiquark ($qq\bar{q}$) known as *baryons* or in bound states of a quark and an antiquark ($q\bar{q}$) known as a *mesons*. These bound states are collectively known as

[9]There are many reasons for this. Part of the difficulty is that quarks are confined to hadrons (Sec. 1.1.1.3). The quark masses are input parameters to the SM they are light (especially u,d, and s) compared the characteric energy scale of QCD.

[10]Collectively known as *flavors*.

[11]A quark has color (RBG), while an antiquark has anticolor ($\bar{R}\bar{B}\bar{G}$).

hadrons and required to be colorless[12]. The top quark, however, is an exception in that it does not form bound states. For a 2-body decay (like the top quark which decays to two other particles, the W boson and the b quark), the decay rate in the decaying particle's rest frame, Γ, is given by the following equation:

$$\Gamma = \frac{|\mathcal{M}|^2}{32\pi^2}\frac{|\vec{p}_1|}{M^2}\, d\phi_1\, d(\cos\theta_1) \tag{1.2}$$

Here $\mathcal{M}$, is an amplitude connecting the probability of a decay to a given final state given a specific initial state[13], $\vec{p}_1$ is the momentum of one of the daughters (or final state) particles, ϕ_1 and θ_1 are angles describing the possible decay direction of the same daughter particle, and M is the mass of the parent (or initial state) particle. The width has an inverse dependence on the mass of the initial particle, but the matrix element must also be taken into account. A heavier particle has more possible decay paths and the general result is a larger width. The lifetime of a partcle is $1/\Gamma$, so generally, heavier particles decay faster than lighter particles. In the case of the top quark, the lifetime is shorter than the effective lifetime of the force that binds quarks together in hadrons, the strong force (discussed later). This makes the top quark a very good choice for probing the properties of quarks.

[12]In fact, they are required to be in a *color singlet*, essentially a superposition of R$\bar{R}$, B$\bar{B}$, and G$\bar{G}$, so they can interact colored gluons.

[13]Feynman diagrams are discussed in Sec 1.1.2.1.

1.1.1.3 Hadrons

Most of the mass of a hadron is not due to the mass of its constituent quarks, but rather due to the kinetic energy of the quarks (they are in motion "inside" the hadron) and Quantum ChromoDynamics (QCD) processes. QCD will be discussed in a later section, but the relevant idea is that there is a sea of energy in which quarks and gluons[14] are constantly created and annihilated.

The two most well-known baryons are the *proton* and *neutron*. The proton (uud) has a mass of 938 MeV, spin of 1/2, and electric charge of +1 (2/3+2/3-1/3). The neutron (udd) has a mass of 940 MeV, spin of 1/2 and no electric charge. These are the only two stable baryons, meaning that any other baryon will eventually decay to them.

Mesons are composed of a quark-antiquark pair. The most famous ones are the J/Ψ ($c\bar{c}$) and the upsilon ($b\bar{b}$). Neither of these bound states fit with the accepted quark model at the time it was discovered. The discovery of the J/Ψ and Υ mesons were used as strong evidence of the existence of charm and bottom quarks, respectively, thereby experimentally confirming expansions to the quark model, which would be the basis for the SM.

[14] Gluons are discussed in Sec. 1.1.2.3.

1.1.2 Bosons

All the observable matter in the universe is composed of the fermions previously discussed. These particles interact through four forces: gravitational[15], weak, electromagnetic, and strong, ordered with increasing strength. These forces are mediated or "carried," by particles known as *bosons*. All bosons have integer spin (0 or 1).

1.1.2.1 Weak Force ($W^{\pm}$,Z^0)

The weakest[16] of the forces included in the SM, the weak force affects all fermions. The weak force is mediated by the $W^{\pm}$ and Z^0 bosons. The $W^{\pm}$ are a particle/antiparticle pair and have a mass of 80.4 GeV while the Z^0 is its own antiparticle and has a mass of 91.2 GeV. All three bosons have a spin of 1. As discussed in Sec. 1.2, the massive nature of these bosons leads to the short lifetime and, by extension, short range of the weak force. The W bosons are responsible for *flavor changing currents*. This is the mechanism by which a quark can "change," into another quark by the emission of a W boson. An example of this is the emission of a electron when a neutron decays to a proton. On a particle physics level, what is happening to the proton is shown by in Fig. 1.2. This kind of diagram is an example of a *Feynman Diagram*.

In classical mechanics, the interaction between objects is definite and well-defined. When a baseball bat connects with a baseball, the momentum and energy transfer

[15]Despite best efforts, gravity is not a part of the Standard Model and is beyond the scope of this work

[16]Specifically with respect to coupling constants. There are times when the weak force will be the dominant force in an interaction.

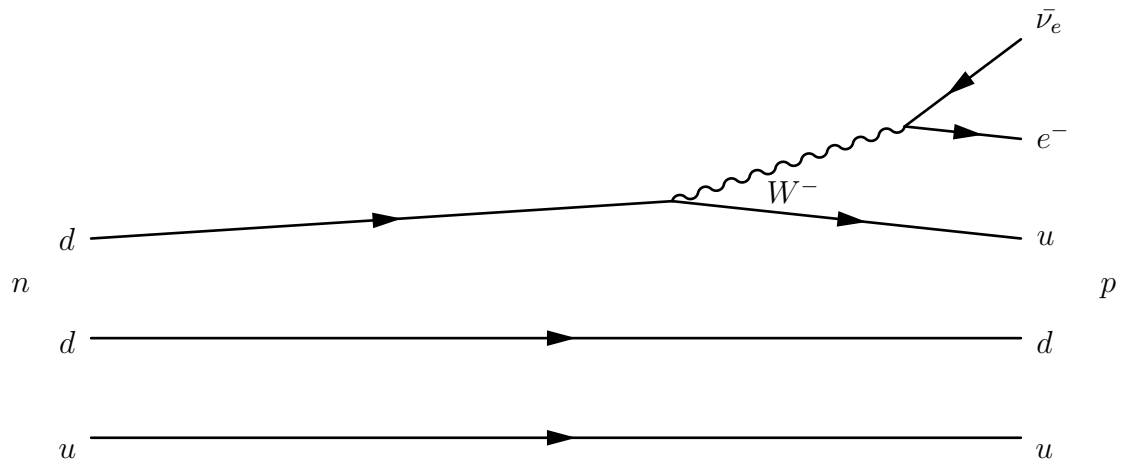

Figure 1.2: β^- decay of the neutron: The decay of a neutron (the 3 bound quarks on the left) to a proton (quarks on right) via the weak force, emitting an electron and an antineutrino

is well-understood and the result can be accurately predicted. In the quantum mechanical analogue, the bat and ball have a probability to interact in several different ways: the ball can be hit as in the classical example, the bat and ball can miss each other, or perhaps the bat is broken by the ball[17]. Quantum mechanically, the result is a superposition of these (and in fact all) possibilities[18]. The probability of each of these outcomes is proportional to a quantum-mechanical amplitude ($\mathcal{M}$), where $\mathcal{M}$ is an infinite series of terms[19]. Richard Feynman devised a pictographic representation for these complex calculations, the eponymous *Feynman diagrams*.

A Feynman diagram is a pictogram that represents a single specific particle interaction[20]. In a Feynman diagram, fermions are represented by solid lines while bosons

[17]These do occur classically, but are not relevant for this example.

[18]In a gross over-simplification, each of the outcomes both happens and does not happen at the same time.

[19]Any number of initial conditions (ie. a slightly changed grip) could still result in a hit.

[20]Each diagram represents just one term in the infinite series of an amplitude.

are represented by wavy lines. The arrows don't indicate movement in a traditional sense, but rather the flow of quantum numbers. For conserved quantities[21], the sum coming into a vertex must equal the sum leaving the vertex. A particle will have an arrow pointing to the right—antiparticles will have arrows pointing to the left. In Fig. 1.2 and all other Feynman diagrams presented, time moves from left to right. In this case, the initial state corresponds to a neutron (bound state of *udd*). At some point in time, one of the down quarks in the neutron emits a W^- boson and becomes an up quark. At that point the hadron is now a proton and doesn't participate any further. The W^- then decays into a e^- and an $\bar{\nu_e}$. Note that the $\bar{\nu_e}$ has an arrow moving to the left while everything else moves to the right. Only the particles on the far left or the far right can be observed—anything else is solely internal. As an infinite number of diagrams can contribute to the same process, it's important to remember that a Feynman diagram only represents one possibility of what might happen. [22,23].

It is not enough to know that a reaction can happen, the rate of the reaction is equally important, especially in particle physics. The rate of weak interactions is influenced by the Cabibbo-Kobayashi-Maskawa (CKM) matrix shown in Fig. 1.3 (Ref. [15]). In general, any of the up-type quarks (uct) can become a down-type (dsb) by the emission or absorption of a W boson and vice versa. However, due to the near diagonal-nature, decays are usually intra-generational. The CKM matrix

[21]Such as electric or strong charge.

[22]Figs. 1.4 and 1.5 both represent possibilities for $e^+e^- \rightarrow e^+e^-$. The actual result is these plus an infinite number of other diagrams.

[23]Though there are an infinite number of possible diagrams, in practice only a few diagrams will significantly contribute to a process. These are termed the *Leading Order* (LO) diagrams.

$$\begin{bmatrix} |V_{ud}| & |V_{us}| & |V_{ub}| \\ |V_{cd}| & |V_{cs}| & |V_{cb}| \\ |V_{td}| & |V_{ts}| & |V_{tb}| \end{bmatrix} = \begin{bmatrix} 0.97427 \pm 0.00014 & 0.22536 \pm 0.00061 & 0.00355 \pm 0.00015 \\ 0.22522 \pm 0.00061 & 0.97343 \pm 0.00015 & 0.0414 \pm 0.0012 \\ 0.00886^{+0.00033}_{-0.00032} & 0.0405^{+0.0011}_{-0.0012} & 0.99914 \pm 0.00005 \end{bmatrix}$$

Figure 1.3: The CKM matrix, representing the amplitude for a flavor-changing weak interaction. The probability of an interaction is given by the amplitude squared, $|V|^2$.

can be written down as a function of 4 independent parameters, which are important inputs to the Standard Model. These values have been found experimentally. It is worth mentioning that Fig. 1.3 isn't the complete picture of the CKM matrix. The true CKM matrix contains a complex phase that allows for CP-violation[24].

Unlinke the W bosons which allow one type of quark to decay into another, the Z^0 always couples to a fermion and its own anti-fermion. In addition, as the Z^0 is uncharged it can interact with neutral fermions such as the neutrino, without the need for a charged lepton to be present.

1.1.2.2 Electromagnetic Force (γ)

All particles with electric charge are capable of interacting through the Electro-Magnetic (EM) force. The EM force is mediated by the massless[25], chargeless, spin-1 photon (γ). Though the EM is the most classical of the forces in the SM, it does have quantum effects governed by Quantum ElectroDynamics (QED). One of the most common interactions, electron-positron annihilation, is shown in Fig. 1.4. In this

[24]C stands for charge conjugation (charge → anticharge). P stands for parity (mirror image reversal). Combined, CP relates matter to antimatter. CP-violation means that matter and antimatter don't behave identically and may explain why the universe is dominated by matter.

[25]Specifically, no rest mass. The photon has a mass due to its momentum.

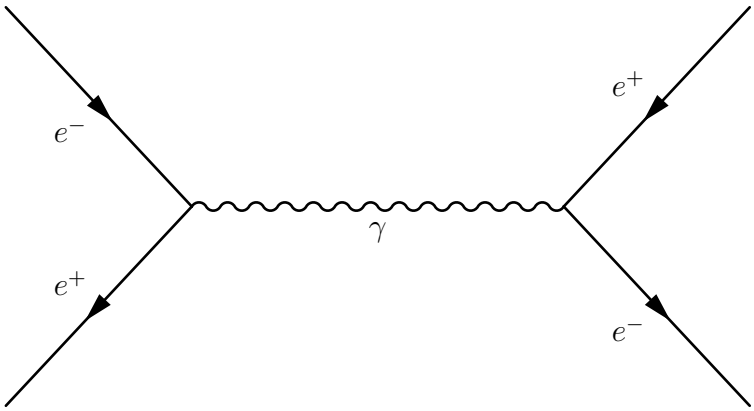

Figure 1.4: Feynman diagram of e^+ e^- annihilation. The incoming particles annihilate and create a photon. The photon can then decay to any allowable state. In this case, it decays back to a e^+-e^- pair.

case, an electron and positron (antielectron) meet and annihilate each other. The energy is created in the form of a photon. This process can also be reversed—the photon can spontaneously decay to an electron-positron pair. In addition to annihilation, photons may also scatter off particles, as shown in Fig. 1.5. As they have charge, quarks can be substituted for the leptons in these examples. Because the γ has zero rest mass, the range of the EM force is infinite.

1.1.2.3 Strong Force (g)

The strongest force in the standard model is appropriately named the strong force. The strong force is mediated by a set of massless, electrically neutral, spin-1 particles known as gluons (g). The strong force only interacts with quarks and other gluons. As discussed in Sec. 1.1.1.2, quarks are confined to bound states of $q\bar{q}$ or (qqq). These quarks are bound together by gluon exchange as in the Feynman diagram in Fig. 1.6.

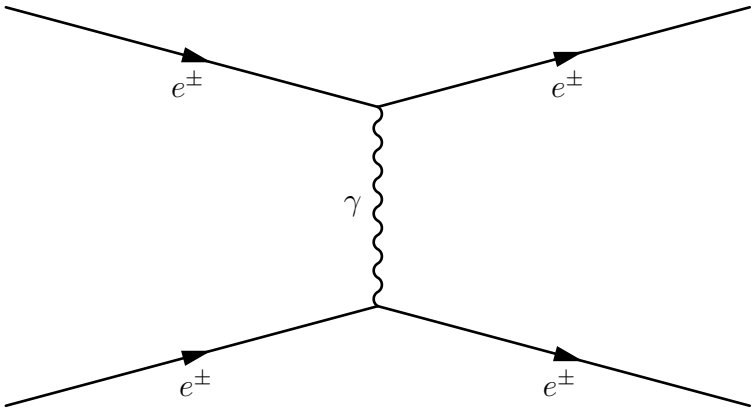

Figure 1.5: Feynman diagram of $e^{\pm}$ scattering. Any combination of e^{+} and e^{-} can interact by exchanging a photon.

As both the gluon and photon are massless, it would seem likely that the strong force should have infinite range. The nature of the strong force precludes this and is what causes quark confinement. The strong force between two quarks can be thought of as a rubberband stretched between a person's hands. As the hands (quarks) are pulled farther apart the resistance (strong force) becomes stronger and stronger. At a certain point, either the rubber band will break or the person will not be able to pull his hands further apart. When this happens in the quantum mechanical strong process, the energy of the system is such that a quark-antiquark pair can be created. Each of the original particle pairs with one of the new particles, exchanging gluons. This can repeat many times and the resulting cascade is called a *jet* in particle physics[26]. Fig. 1.7 demonstrates what happens if the quarks in Fig. 1.6 travel apart and hadronize in this manner.

The concept of color charge was discussed in Sec. 1.1.1.2 and the three types

[26]Jets and their reconstruction will be discussed in Ch. 2

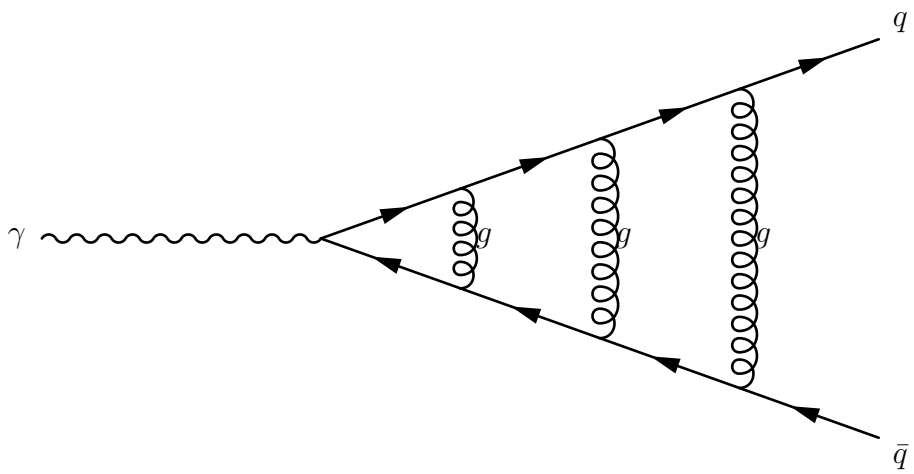

Figure 1.6: Feynman diagram of photon decay to $q\bar{q}$ pair. The quarks can exchange any number of gluons.

of color charge were introduced: Red (R), Blue (B), and Green (G). All observable particles must be colorless, with the following possibilities: (RBG), $(\bar{R}\bar{B}\bar{G})$, $(R\bar{R})$, $(B\bar{B})$, and $(G\bar{G})$. Gluons can carry net color, such as $(R\bar{G})$, but they are confined to a hadron and therefore not directly observable. This is a simplification and in fact, there are 8 possible linearly independent color combinations, which yield 8 distinct gluons. An example set is:

$$(R\bar{B}+B\bar{R})/\sqrt{2} \qquad (R\bar{G}+G\bar{R})/\sqrt{2} \qquad (B\bar{G}+G\bar{B})/\sqrt{2}$$
$$-i(R\bar{B}-B\bar{R})/\sqrt{2} \qquad -i(R\bar{G}-G\bar{R})/\sqrt{2} \qquad -i(B\bar{G}-G\bar{B})/\sqrt{2}$$
$$(R\bar{R}-B\bar{B})/\sqrt{2} \qquad (R\bar{R}+B\bar{B}-2G\bar{G})/\sqrt{6}$$

As mentioned previously, hadrons are in a color singlet:

$$(R\bar{R}+B\bar{B}+G\bar{G})/\sqrt{3}$$

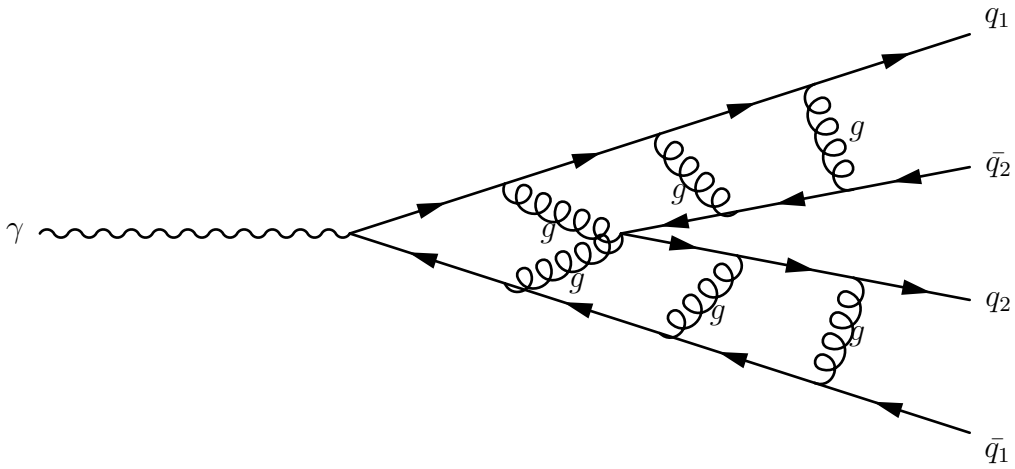

Figure 1.7: Feynman diagram of photon decay to $q\bar{q}$ pair which then hadronizes. As time passes, the original quark-antiquark pair, (q_1,$\bar{q}_1$) move apart and a new quark-antiquark pair is created (q_2,$\bar{q}_2$). This is the idea behind jets.

1.1.3 The Higgs Boson (H)

The last fundamental particle in the Standard Model is also the most recently discovered. The Higgs boson (H) was discovered at the European Organization for Nuclear Research (CERN) in 2012. The discovery was long-anticipated and the 2013 Nobel Prize in Physics was awarded to Higgs and Englert[27] who had theorized the particle in 1964. The Higgs boson has a mass of 125 GeV, no electric or color charge, and spin of 0.

In the simplest interpretation, the Higgs boson is responsible for other particles having mass. More precisely, the Higgs boson is the result of a Higgs field that permeates all space. The field can be thought of as a room full of people, milling about. Then a famous person (particle) enters the room. As the person passes

[27]Englert had a co-author on the awarded work, Robert Brout, who died in 2011. Due to the Nobel Prize committee's policy of not posthumously awarding a Prize, Brout was ineligible.

through the room, people (Higgs field) tend to clump around the famous person as she moves through the room. The more famous the person, the more the person gets stopped by the people in the room and the slower she moves through the room. In the same way, a particle that has a stronger interaction with the Higgs field will have more mass than one which interacts less.

The top quark is the heaviest quark and therefore couples the most strongly to the Higgs boson. As the Higgs mechanism isn't well understood, studying the properties of the top quark is an important way to further this understanding.

1.2 Parton Distribution Functions

Hadrons are made up of quarks and gluons, collectively called *partons*. Rather than being made up of two up quarks and a down quark, the proton (and other hadrons) is made of many different partons, governed by a *Parton Distribution Function* (PDF). A PDF is generally presented as the momentum fraction of a parton[28] (x) vs the momentum fraction times the distribution function ($x \times f(x)$). One set of PDFs in use is shown in Fig. 1.8. By far, the most likely parton to be found in the proton is the gluon. However, these tend to carry less of the overall momentum (low x).

[28] More specifically, x is the fraction of the proton's momentum carried by a given parton.

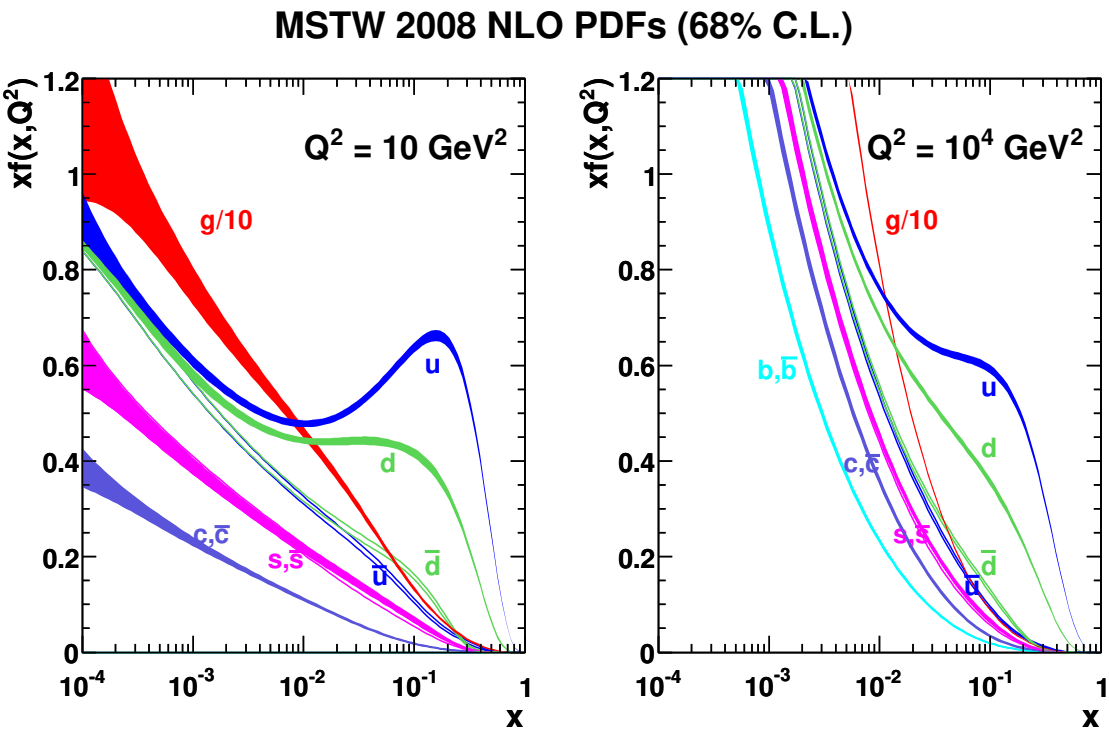

Figure 1.8: Parton Distribution Function for the proton at the LHC, plotted as the momentum fraction (x), vs the distribution function times x, ($x \times f(x)$). Image taken from Ref. [2].

1.3 Cross Sections

The chance of hitting a fixed-sized target with no long range interaction is directly proportional to the geometric cross section of the target. In particle physics, the *cross section*, σ, corresponds to how likely an interaction is to take place and is given by Eq. 1.3.

$$\sigma \equiv \frac{N}{\epsilon A \mathcal{L}} \tag{1.3}$$

Here N is the number of times a given process was found (for example, $t\bar{t}$) and $\mathcal{L}$ is the luminosity[29] collected, ϵ is the reconstruction efficiency and A is the geometric acceptance of the detector. The parameters ϵ and A account for imperfections in the experimental setup. Several theoretical cross sections can be seen in Fig. 1.9.

The Standard Model predicts a cross section for the production of every particle discussed in this document. Therefore, an accurate cross section measurement serves as a test of the SM. Of special interest is the $t\bar{t}$ production cross section because the Large Hadron Collider (LHC) produces more than 30 times as many top quarks than have ever been produced before, providing the best opportunity to measure top quark properties. A measurement that deviates from the SM prediction would indicate new physics. In fact, $t\bar{t}$ is an important background in many searches for new physics, so a measurement inline with the SM would be useful in these searches.

[29]Luminosity is a measure of how many collisions have taken place per unit time per unit area. More accurately, it is the flux of particles through the LHC.

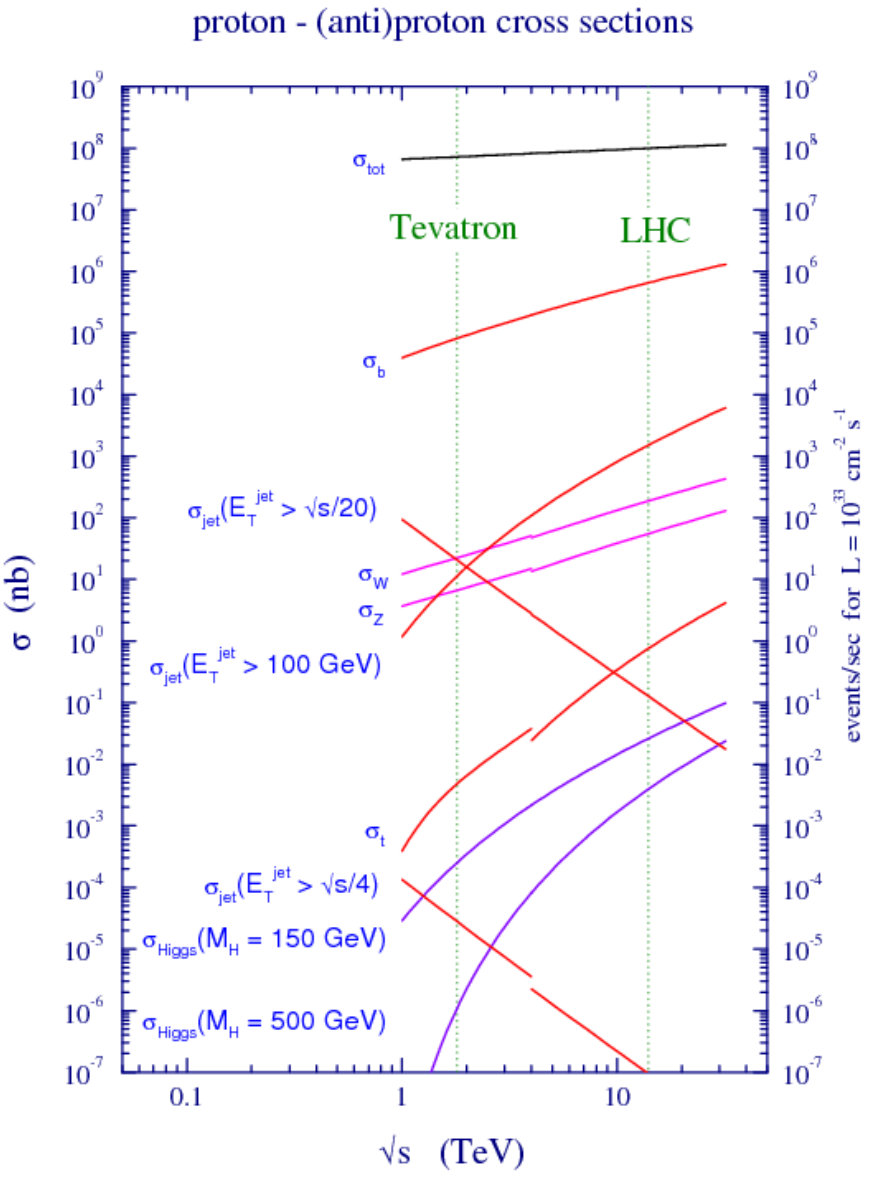

Figure 1.9: Theoretical prediction for scattering cross sections of at the Tevatron and the LHC as a function of center of mass energy, $\sqrt{s}$. Image taken from Ref. [3].

The cross sections shown in Fig. 1.9 are plotted as a smooth function of $\sqrt{s}$. The theoretical cross section for any process can be plotted as a smooth function of almost any parameter, such as mass or energy. However, in practice, only a finite number of interactions occur at the LHC, and fewer still occur that pass all quality and selection cuts. This lack of data severely limits the granularity or "smoothness" of the cross section that can be reconstructed. In this case, it may be more instructive to look at the *differential cross section* instead. For example, the differential $t\bar{t}$ cross section as a function of transverse momentum[30], $\frac{d\sigma}{dp_T}$, is given by Eq. 1.4:

$$\frac{d\sigma}{dp_T} = \frac{1}{\Delta p_T} \frac{N}{\epsilon A \mathcal{L}} \tag{1.4}$$

where Δp_T is a measure of how finely the measurement[31] is divided.

A Feynman diagram for the hadronic $t\bar{t}$ decay is shown in Fig. 1.10. The W bosons involved in the $t\bar{t}$ decay can themselves decay to leptons instead of quarks. If one W boson does so, the result is called *semileptonic*[32]. If both W bosons decay leptonically, the resulting process is called *leptonic*[33]. The ratio of the cross section for a specific process to the total cross section is called the *branching fraction* (BF). For the $t\bar{t}$ decay, the hadronic branching fraction is:

$$BR = \frac{\sigma_{t\bar{t}}^{had}}{\sigma_{t\bar{t}}^{tot}} = \frac{6}{9} \times \frac{6}{9} = \frac{36}{81} \tag{1.5}$$

[30] Otherwise known as p_T. p_T will be discussed more in Chapter 2.
[31] With respect to p_T in this case
[32] Alternatively muon+jets.
[33] Or dimuon.

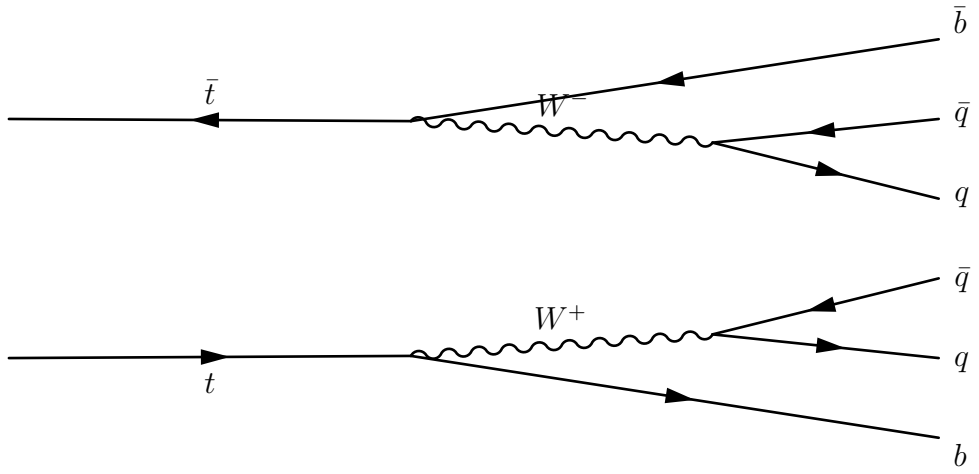

Figure 1.10: Feynman diagram of hadronic $t\bar{t}$ decay.

The value for this branching fraction is found by considering the possible decays. A top quark always decays to a W boson and a bottom quark[34]. The W can then decay to any quark-antiquark pair (6 options), or to any lepton-antineutrino pair (3 options), for a hadronic probability of 6/9. The second W has the same probabilities, for a total branching fraction of 36/81.

The difference between the hadronic $t\bar{t}$ cross section and the more general[35] $t\bar{t}$ cross section measured in the hadronic channel is this branching fraction.

1.4 Differential $t\bar{t}$ Cross Section

At this point, the focus of this thesis can be pretty well defined. The $t\bar{t}$ cross section has been predicted by theorists and their results used to generate Monte

[34]From the CKM matrix (Fig. 1.3), the matrix element, $|V_{tb}| = 0.9914$. The probability of this decay is $|V_{tb}|^2$, or $> 99.8\%$ of the time.

[35]And theoretically relevant.

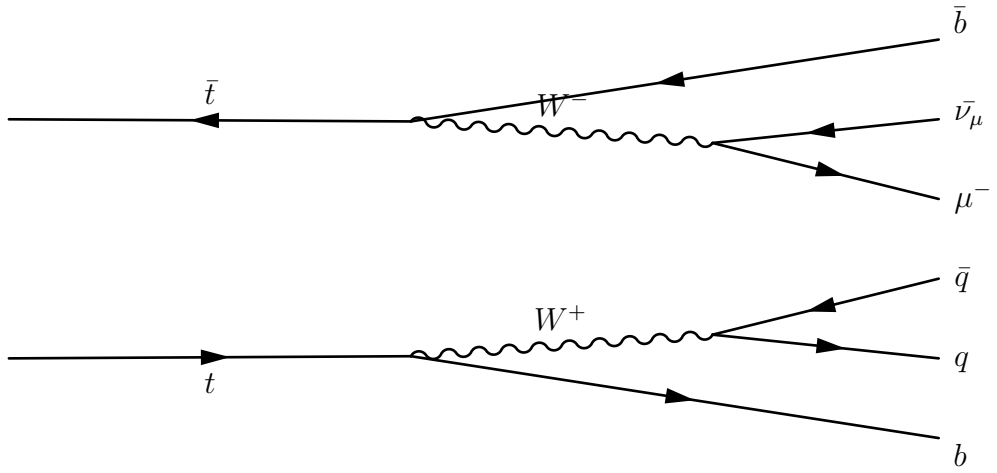

Figure 1.11: Feynman diagram of semileptonic (muon + jets) $t\bar{t}$ decay.

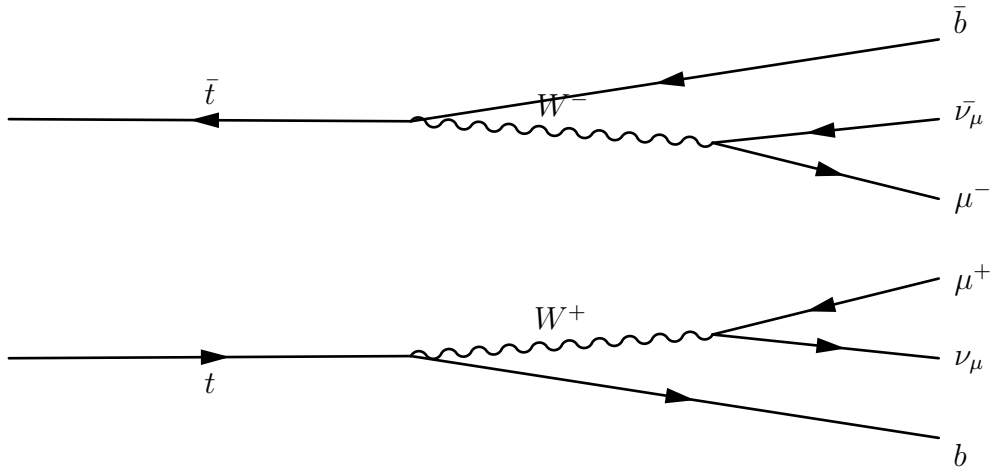

Figure 1.12: Feynman diagram of fully leptonic (dimuon) $t\bar{t}$ decay.

Carlo $t\bar{t}$ samples. The differential cross section for such a distribution is shown in Fig 1.13.

Theorists have taken all the Feynman diagrams for potential $t\bar{t}$ production and summed them to come up with the inclusive $t\bar{t}$ cross section. The Feynman diagrams which contribute the most to the a cross section are called the *leading order* (LO) diagrams. The LO diagrams for $t\bar{t}$ production are shown in Fig. 1.14. Diagrams (b), (c), and (d) are the dominant contributions at the LHC (Large Hadron Collider) due to the prevalance of gluons in the proton pdf[36].

Each of the LO diagrams has two vertices (quark-gluon or gluon-gluon interaction points). However, due to the nature of the strong force, "extra" gluons can be added to these diagrams without changing the initial or final state. These constitute the *Next-to-Leading Order* (NLO) diagrams, many of which are shown in Fig. 1.16.

An additional component of the NLO cross section comes from *Initial State Radiation* (ISR) or *Final State Radiation* (FSR). Figure 1.15 shows two NLO constributions to the $t\bar{t}$ cross section from radiation. The full NLO cross section includes similar diagrams for each of the processes in 1.14. A gluon will hadronize like a quark, so ISR and FSR show up as an additional hadronic jet. As the results presented do not require exactly 4 jets[37], these processes also contribute to the measured cross section.

The next chapter will discuss the experiment and detector used to make this measurement.

[36]Figure 1.8.

[37]Each top decays to a W which then decays to a q q' pair, for a total of hadronic jets.

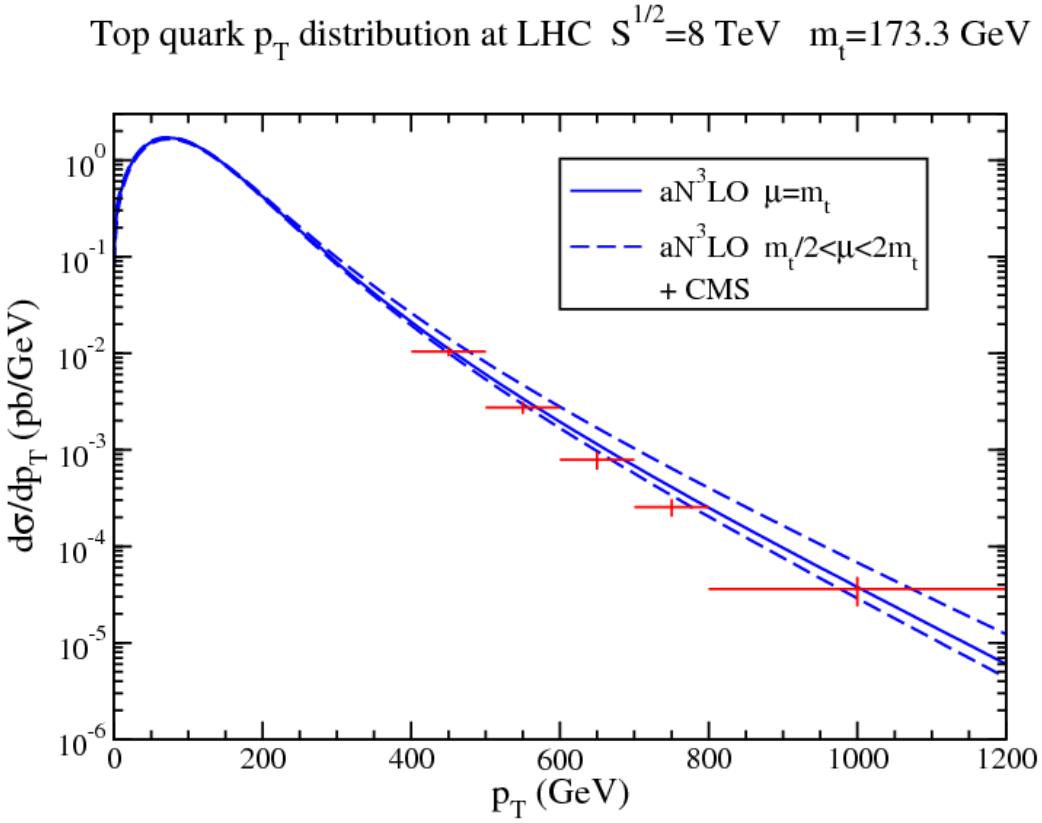

Figure 1.13: Theoretical differential top quark cross section (blue) with CMS semileptonic data (red). Image taken from Ref. [4].

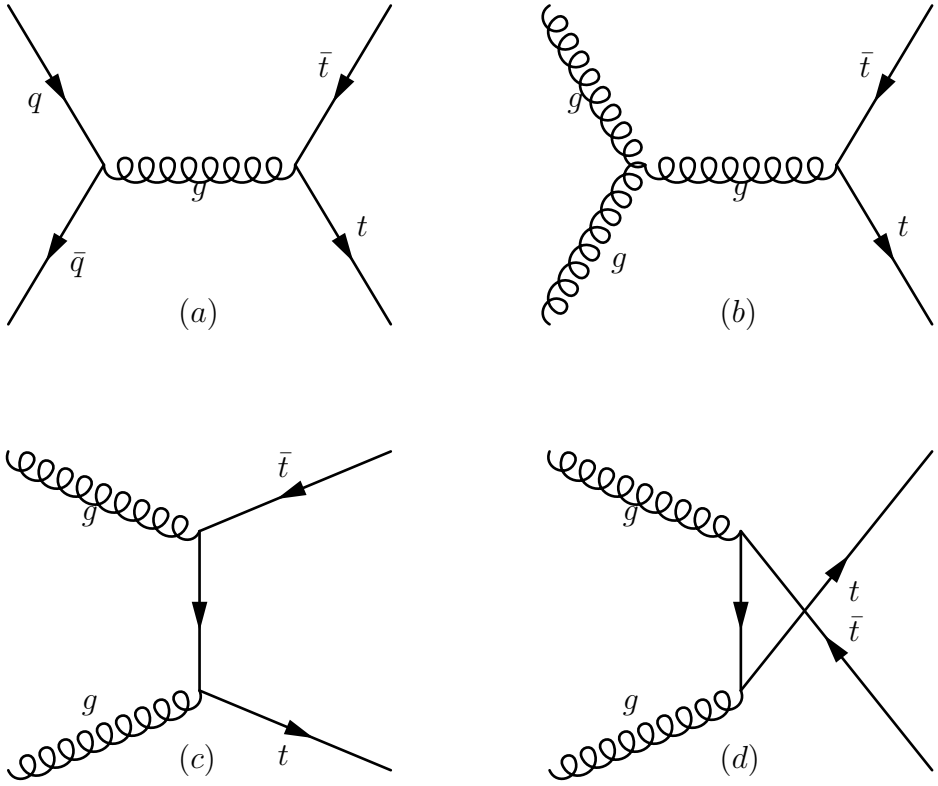

Figure 1.14: Feynman diagrams of LO $t\bar{t}$ production.

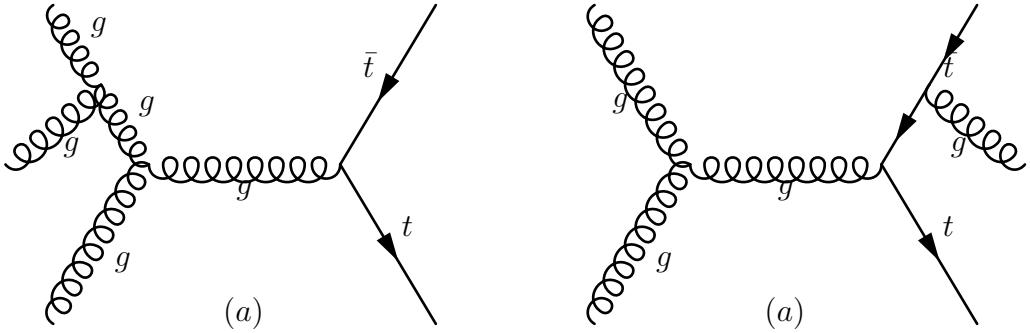

Figure 1.15: Feynman diagrams of NLO ISR (a) and FSR (b) $t\bar{t}$ production. The extra gluon can be on either the top of bottom leg of the diagram.

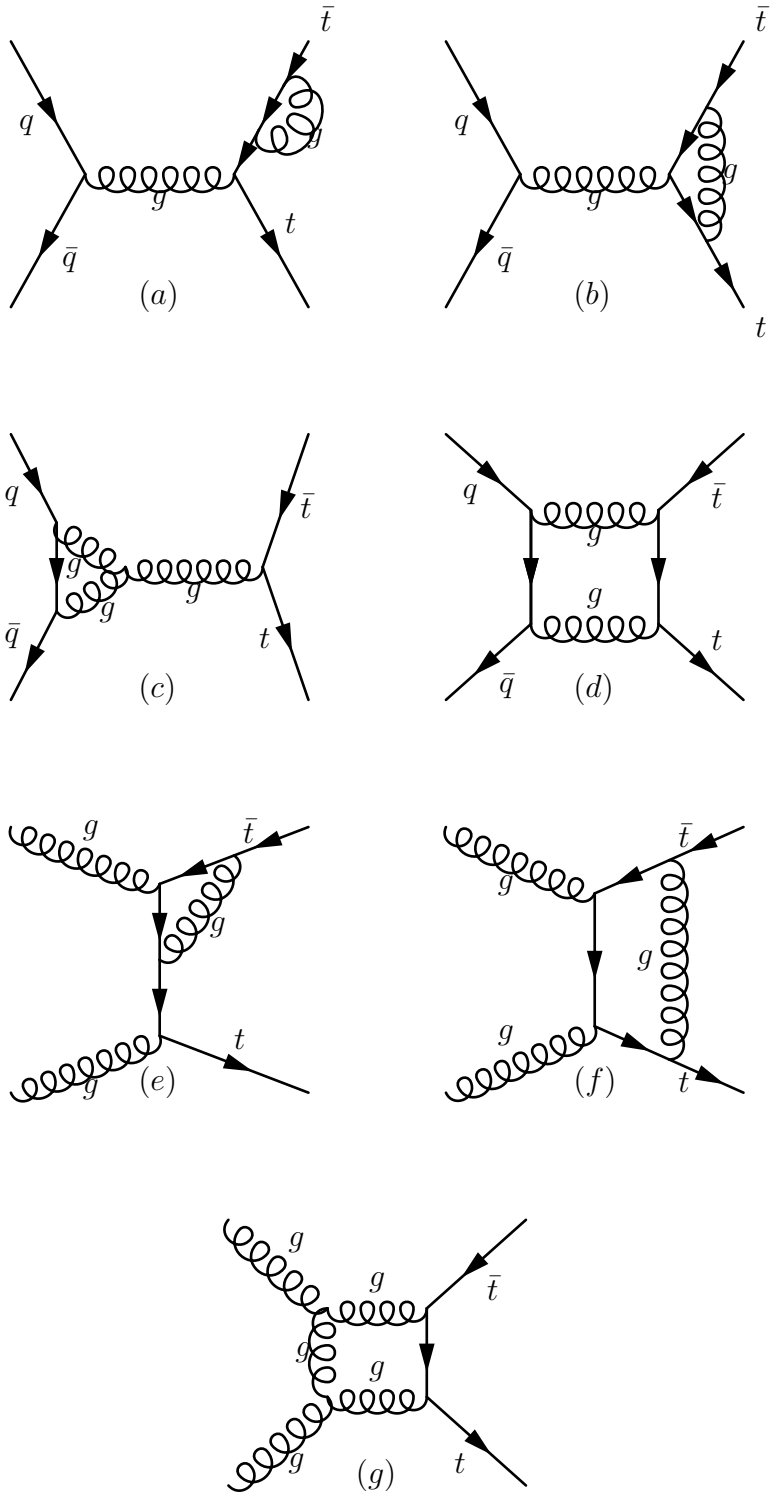

Figure 1.16: Feynman diagrams of NLO $t\bar{t}$ production.

Chapter 2

Experimental Setup

2.1 The LHC at CERN

The Large Hadron Collider (LHC) located at CERN, straddling the Swiss-French border, collides protons at some of the highest energies available to mankind. As seen in Fig. 1.9, the production cross-section for many processes increases with increasing energy. At particle accelerators, the energy is quantified as the center-of-mass energy ($\sqrt{s}$). The energy of the output particle would be $\sqrt{s}$ if the protons were completely annihilated. In practice, this is never the case as protons are composite particles—the particles created in the interaction have a fraction of $\sqrt{s}$ energy. The accelerator is scheduled for periods of proton-proton collisions separated by periods where the accelerator is shutdown for maintainance and upgrades. Run I consists of data taken

between 2009 and 2013. During this period 19.7 fb^{-1}[1,2] of data was collected at $\sqrt{s} = 8$ TeV. Data-taking for Run II is ongoing at $\sqrt{s} = 13$ TeV and expected to last until 2018[3]

The LHC (Fig. 2.1) is a ring[4] 27 km in circumference used for proton-proton collisions. Hydrogen gas is ionized, stripping the electrons, to get protons which go through a series of accelerators until they are injected into the LHC as two separate beams. Once inside the LHC, the two beams of protons are accelerated in opposite directions until they each reach 4 TeV and then collided every 50 ns. The LHC has four experiments[5] located at collision points around the ring to collect data from these collisions.

2.2 Compact Muon Solenoid

The Compact Muon Solenoid (CMS) (Fig. 2.2[6]) is one of the two all-purpose detectors collecting data from the LHC. CMS was designed to have as close to 4π coverage around the interaction point as possible, in an attempt to collect all the par-

[1] A unit of area. 1 fb = 10^{-15} barn. 1 barn = 10^{-28} m^2.

[2] As seen in Eq. 1.3, cross section and luminosity are inversely related. Cross section has units of area (fb), so Luminosity has units of inverse area (fb^{-1}).

[3] This document presents results from Run I, so all future discussions will assume $\sqrt{s} = 8$ TeV.

[4] Called a synchrotron.

[5] These are the Compact Muon Solenoid (CMS) detector, A Large Ion Collider Experiment (ALICE), A Toroidal LHC ApparatuS (ATLAS), and the Large Hadron Collider beauty (LHCb) experiment.

[6] Including a bonus picture of the author!

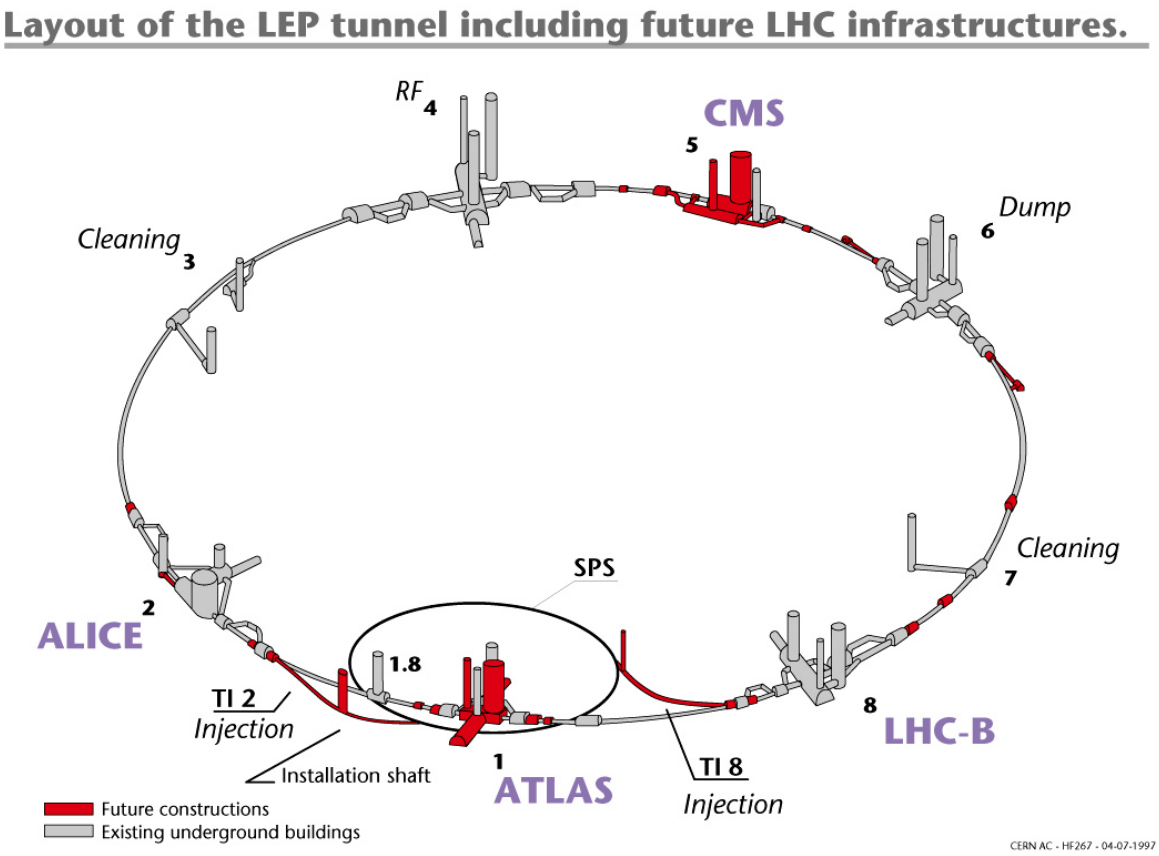

Figure 2.1: A schematic of the Large Hadron Collider showing the eight potential interaction points and the four detectors occupying half of them: CMS, ALICE, ATLAS, and LHC-B. Image taken from Ref. [5].

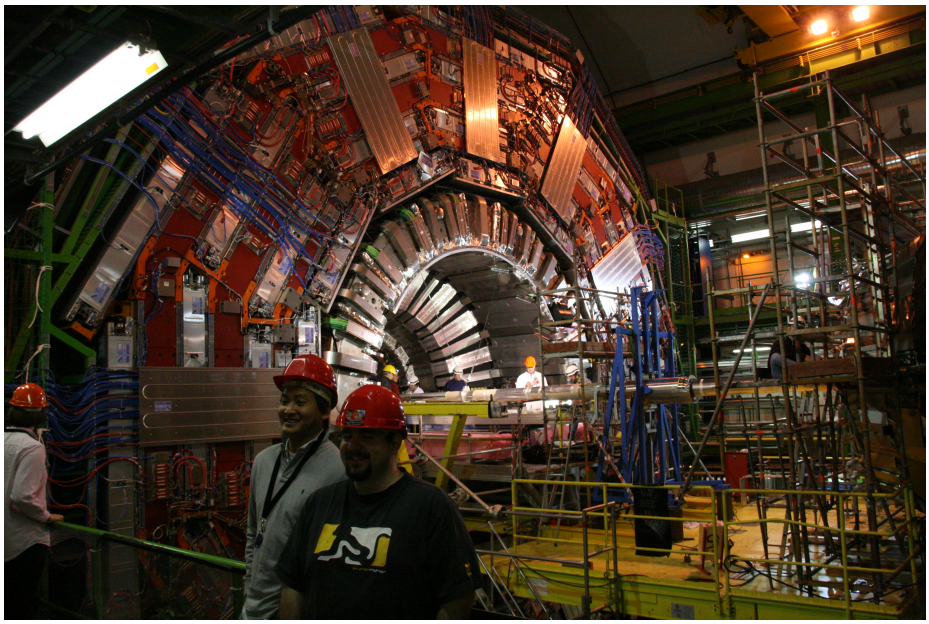

Figure 2.2: Picture of the Compact Muon Solenoid detector. Pictured in foreground: Nhan Tran (left) and Dave Fehling (right).

ticles created during the collision. Many different types particles are created[7], including electrons, muons, photons, bosons, and hadrons. As previously discussed,[8] these particles can interact in many different ways necessitating many different detector systems acting in concert to accurately reconstruct the collision. A cross sectional-view of the CMS detector is shown in Fig. 2.3. Much more detail about the following subdetectors can be found in the CMS technical design report (Ref. [16]).

[7]Fig. 1.8 demonstrates that different partons can collide and Fig. 1.9 gives an example of several different processes that can occur after the collision

[8]Sec. 1.1.2.

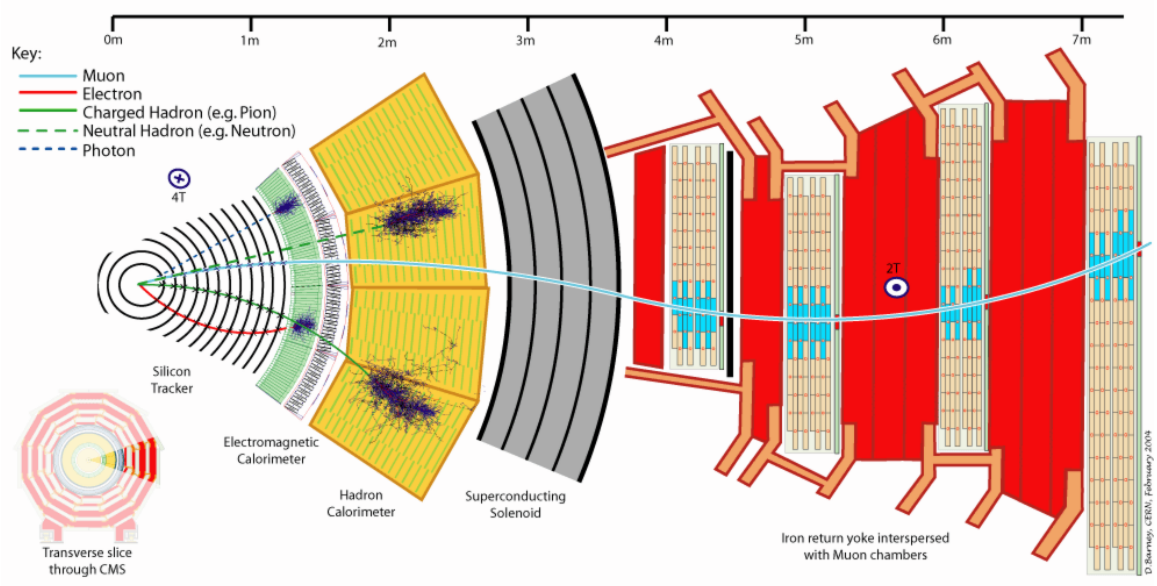

Figure 2.3: Cross-sectional view of the Compact Muon Solenoid detector, showing how different particles interact with the individual subdetectors. Image taken from Ref. [6].

2.2.1 Magnet

A superconducting solenoid[9] lies at the heart of CMS[10]. The 3.8 T[11] magnetic field created by the magnet is very useful in helping other parts of the dectector accurately identify particles. Charged particles exposed to magnetic fields travel in curves instead of straight lines; the direction of the curve allows positive and negative electric charge to be distinguished. The amount the path curves depends on both the particle's mass and velocity, so the magnetic field allows the measurement of the

[9] Also known as an electromagnet.

[10] Both literally and figuratively.

[11] A unit of magnetic field strength. 1 T = 1 $\frac{kg}{A*s^2}$ where kg is the SI unit of mass and A (Ampere) is the SI unit for current. The Earth's magentic field is $3.2 \times 10^{-5} T$.)

tranverse momentum[12] (p_T) of charged particles.

2.2.2 Silicon Tracker

The first part of CMS that a particle passing through the detector will encounter is the silicon tracker (Fig. 2.4) broken into four parts: the Tracker Inner Barrel (TIB), the Tracker Outer Barrel (TOB), the Tracker Inner Disks (TID), and the Tracker EndCaps (TEC). In total, the tracking system consists of 75 million separate read-out channels, providing spatial resolution from approximately 50μm to as fine as 10μm. The TIB is made of more than 65 million silicon pixels (pictured in Fig. 2.5). Each pixel detector unit in the TIB contains 16 readout chips (ROCs) bump-bonded to more than 65,000 pixel sensors[13]. The pixels are 100μm x 150μm silicon wafers read out by the ROC in the 3rd dimension by the application of a bias voltage. As a charged particle (such as a jet from a top quark decay) passes through the pixels, it creates electron-hole pairs in the depleted region of the silicon. The freed electrons and holes become separated by the bias voltage and the signal is recorded as a hit. Neutral particles such as the photon or neutron have no ionizing effect and, as such, are invisible to the tracker. The three remaining parts of the tracker are made up of about 10 million silicon strips. The strips vary in size based on their location, but they behave the same way as the pixels. The biggest difference between the two

[12]This will be discussed in Sec. 2.2.2 on the tracker, but many (and in fact most) particles travel at least somewhat in the direction of the beamline. p_T measures the amount of momentum that does not, but rather travels "outward" through CMS.

[13]Each ROC has $80x52 = 4160$ pixel sensors.

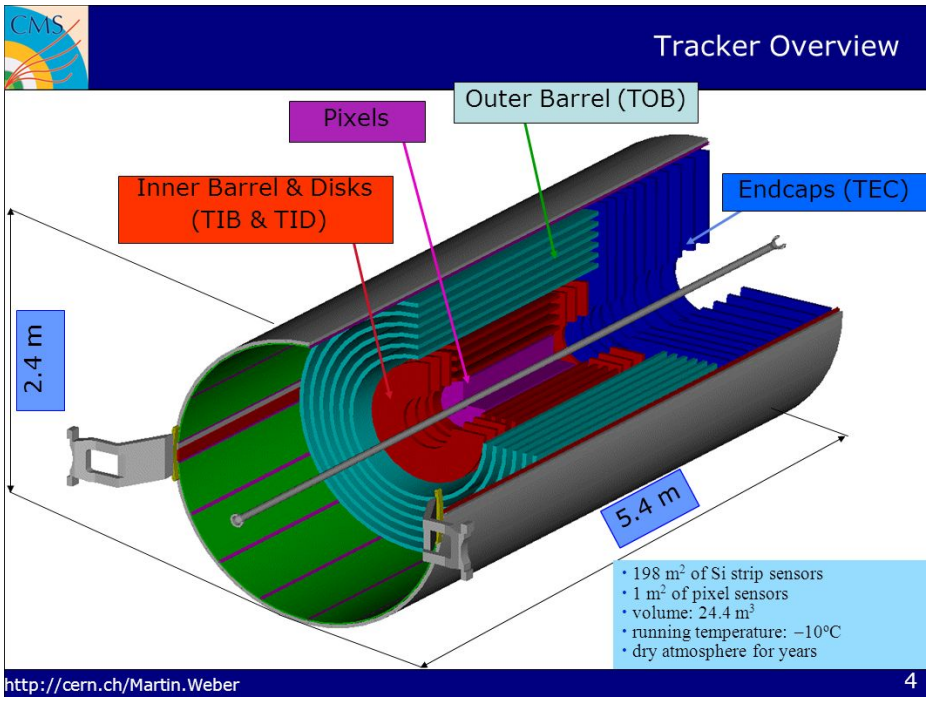

Figure 2.4: A rendering of the CMS tracker with the four parts TIB, TOB, TID, and TEC clearly labelled. Image taken from Ref. [7].

systems is that the pixel system is able to provide a 3-dimensional position, while strips can generally only provide 2-dimensional information. A more in-depth look at CMS tracking and reconstruction can be found in Ref. [17].

As events come from the collision of two equally energetic protons, many created particles are produced nearly at rest and pass through the TIB and TOB. The TIB consists of 3 layers of pixel sensors set 4cm, 7cm, and 11cm from the beamline and is instrumental to reconstructing weak decays[14]. In many other cases, one parton in the collision will have much more energy than the other and the created particles will

[14]The typical lifetime of a b-containing hadron is around 1.5ps. These particles are light and move at nearly the speed of light and travel approximately 450μm before decaying. This is detectable with a resolution of 10μm.

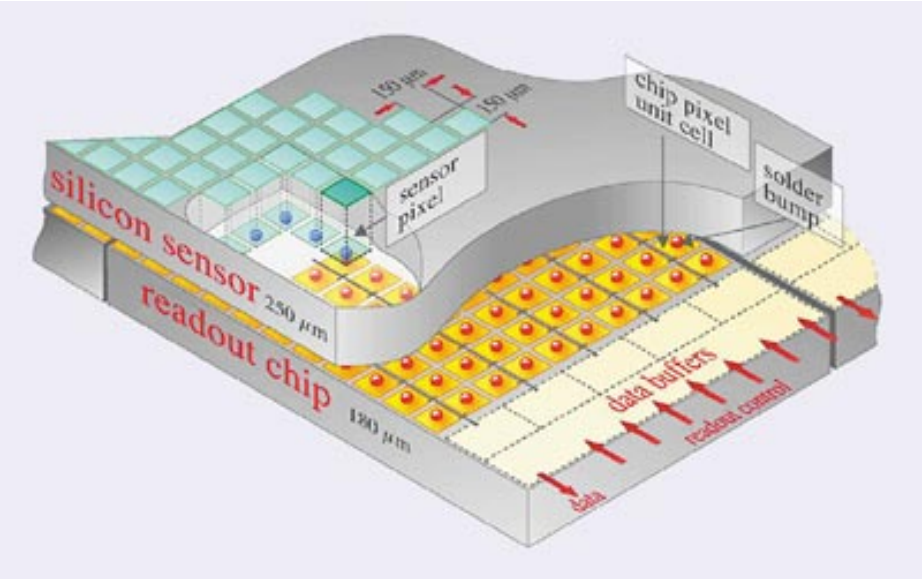

Figure 2.5: A rendering of a silicon pixel detector unit, consisting of a readout chip bonded to silicon pixels. Image taken from Ref. [8].

travel along the beamline. The TID and TEC detectors provide tracking in this case, specifically the region with for $1.0 < |\eta| < 2.5$. η is called the pseudorapidity and is defined as:

$$\eta \equiv -\ln\left[\tan\left(\frac{\theta}{2}\right)\right] \tag{2.1}$$

where θ is the angle of the particle's path with respect to the beamline. This is preferred to θ because differences in η (useful for quantifying the spread of a jet) are lorentz-invariant. This removes the problems created by collisions of partons with mis-matched energy. The tracking detector allows CMS to reconstruct the trajectory of a charged particle by providing up to 15 separate readings of its path.

2.2.3 Electromagnetic Calorimeter (ECAL)

After a particle (charged or uncharged) passes through the silicon tracker, it encounters the Electromagnetic Calorimeter (ECAL) shown in Fig. 2.6. The ECAL is composed of more than 75,000 lead tungstate ($PbWO_4$) crystals. When an energetic electron enters the ECAL, it gives off energy by *bremsstrahlung*: the electron interacts with the electromagnetic field of the lead tungstate atoms, slowing down; this decrease in electron energy is emitted as a photon. As the electron traverses the length of the crystal, it repeatedly interacts with atoms and radiates photons until it has no more energy to do so. Through measuring the light produced during this process, the ECAL is able to measure the energy of the incident electron.

The Feynman diagram in Fig. 1.4 shows electron-positron annhilation to produce a photon. The reverse process is possible where a photon spontaneously becomes an e^+-e^- pair. The ECAL makes explicit use of this to measure the energy of photons. More specifically, the ECAL does not differentiate between electrons and photons. Photons that enter become electron-positron pairs that each release photons through bremsstrahlung. These photons can either be collected to measure the energy or themselves convert to electrons and positrons, and so on until all energy is collected.

Other particles that enter the ECAL are only slightly affected by bremsstrahlung and react minimally with the ECAL.

Figure 2.6: A picture of the Electromagnetic Calorimeter (ECAL) taken during construction. Image taken from Ref. [9].

2.2.4 Hadron Calorimeter (HCAL)

A particle passing through CMS will next encounter the Hadron Calorimeter (HCAL). The HCAL (Fig. 2.7) is made of alternating layers of brass and plastic. Similarly to the ECAL, the HCAL collects photons resulting from hadron showers to measure the energy of an incident hadron. Aside from this, the two calorimeters are quite different. Brass is a fairly dense material[15] and a particle passing through the alloy will eventually collide with an atomic nucleus. This collision results in a shower of many different particles including hadrons, pions[16], electrons and photons. The

[15]As an alloy, the density of brass can vary, but it is around 8.5 g/cm^3, or about 75% of the density of lead.

[16]The lightest mesons, made of combinations of $(q_1\bar{q_2})$ where $q1$ and $q2$ are either up or down quarks.

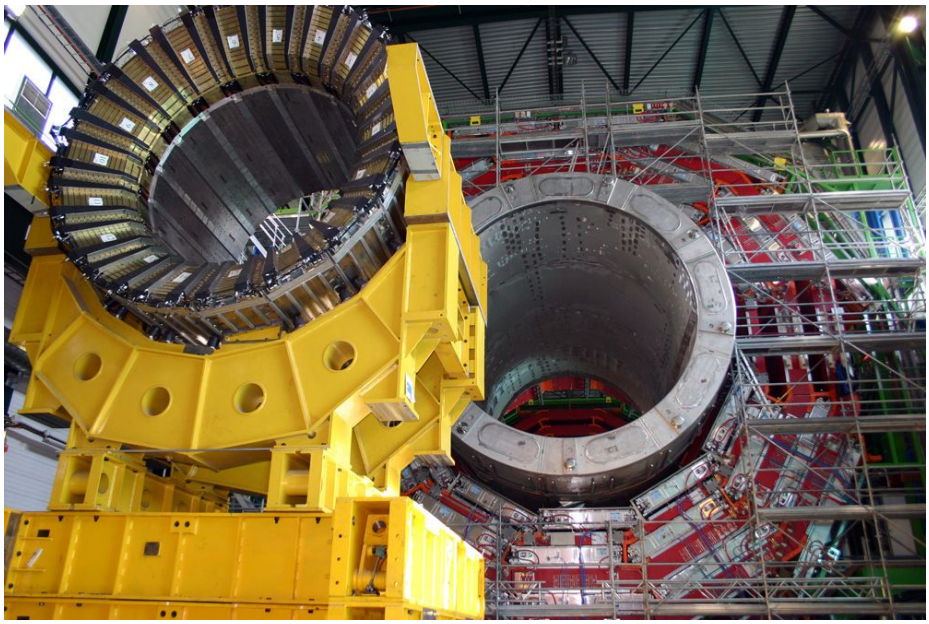

Figure 2.7: A picture of the Hadron Calorimeter (HCAL). Also visible in the background is the muon system and magnet. Image taken from Ref. [10].

charged particles directly ionize the plastic scintillator in the HCAL while the neutral π^0 decays to an e^+-e^- pair which is also collected in the scintillator as it showers electromagnetically. The hadrons continue to hit atomic nuclei, creating these showers, until the particles created don't have enough energy to create a new shower. Unlike the ECAL, the HCAL only collects a small sample of the hadron's energy. Calculating the energy of a hadron in this manner relies on studies relating the energy of emitted photons to that of incident particles. This is an iterative process and data from the LHC is used to refine this *response*. A top jet would create a hadron shower in the HCAL and travel no further. More information about the interactions which govern both the ECAL and the HCAL can be found in Ref. [18].

2.2.5 Muon System

At this point, generally two particles are left that haven't been stopped and measured by CMS. The first is the muon, the heavier[17] version of the election. The muon is somewhat unique in that it is massive enough to not interact with the ECAL, but light enough to pass through the HCAL. In general, muons are useful because, as with any unstable particle[18] they may give insight into physics not normally accessible. Furthermore, muons are easily identified and can therefore be clearly distinguished from background events.

The muon system is divided into two parts: the barrel region (Fig. 2.8) and the endcap region. The barrel muon system is made of three concentric sets of drift tubes; each set contains 4 drift tubes. Each drift tube consists of a charged wire in an Ar/CO_2 gas. Muons (or any charged particle) pass through the gas and ionize it in much the same way the silicon is ionized by a charged particle in the tracker. The electrons thereby created drift to the wire where they are read out and the muon's position can be calculated. The muon chambers are interspersed with the iron return yoke. The magnetic field created by the magnet exists both within and outside the magnet. The return yoke serves to guide the outer magnetic field, giving a more uniform magentic field. The outer magnetic field, however, is oriented opposite to the internal magenetic field—this can be seen in the muon in Fig. 2.3 and must be taken into account. The muon, as a charged particle, travels in an arc because of

[17] About 200x.
[18] Such as the top quark!

the interior magnetic field. However, once the muon has passed the magnet and is traversing the return yokes, it curves in the opposite direction. Some work must be done to match the track the muon makes through the silicon tracker to the track it makes through the muon system.

The endcap muon system uses Cathode Strip Chambers (CSC) instead of drift tubes. CSCs operate much like drift tubes, but are more robust. The CSC consists of a series of positively charged wires in an $Ar/CO_2/CF_4$ gas. A series of negatively charged strips are perpendicular to the wires. When a charged particle passes through the CSC, the gas is again ionized and the electrons are collected by the wires, but the positively charged ions are also collected by the strips. This enables a 2-dimensional position measurement.

Most muons are not stopped in the muon system. Instead, their p_T is measured using information from both the tracker and the muon systems and their energy calculated from this.

2.2.6 Missing Tranverse Energy (MET)

There is one particle that hasn't be detected by any system thus far discussed: the neutrino. In fact, CMS does not have any system in place capable of detecting neutrinos. CMS instead relies on each of the just-discussed detectors to accurately reconstruct everything except neutrions. If, when adding up the energy, there is an

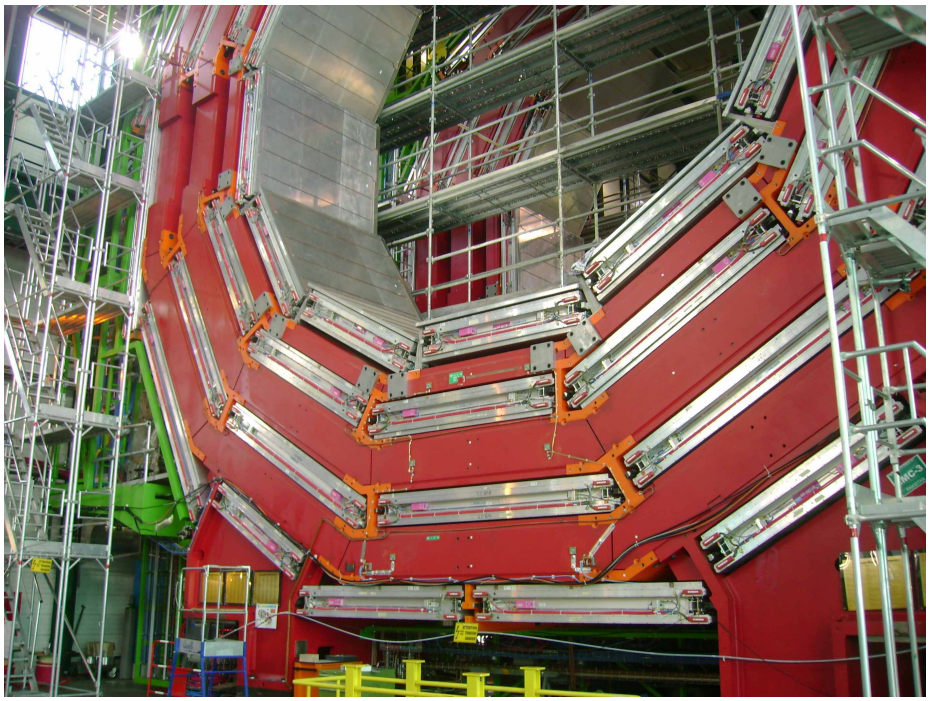

Figure 2.8: A picture of the CMS muon system. The system consists of drift chambers (the silver, metallic sections) interspersed with the iron return yoke(red). Image taken from Ref. [11].

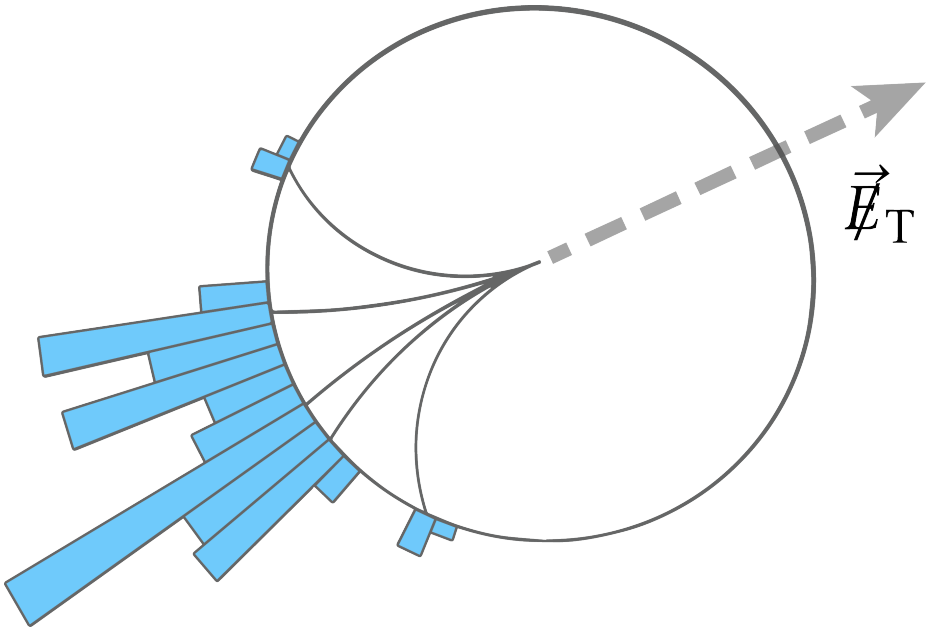

Figure 2.9: Schematic of Missing E_T (alternatively known as MET or $\not E_T$). Image taken from Ref. [12].

imbalance[19] the energy deficit is called *missing transverse energy* (MET[20]). This MET is usually assumed to be a neutrino. However, it could also be due to new physics or a mismeasurement somewhere in the detector.

2.2.7 Trigger

Collisions at the LHC occur every 25-50ns. This corresponds to 20-40 million events every second. It would be impossible to record every one of these events[21],

[19] Because the colliding partons have no transverse energy, conservation of energy states that the sum of E_T in any hemisphere of the detector should be matched by the sum in the other hemisphere.

[20] Alternatively $\not E_T$. On the left hand side, the energy of various particles has been collected. There should be an approximately equivalent amount on the right side—this discrepancy is called MET.

[21] It would also be unnecessary because many events contain no useful or exciting physics.

so CMS uses a series of triggers to identify potentially interesting events and record them. CMS performs a very rough (but quick) reconstruction of an event and looks for very specific signals such as a high-energy jet or a muon—the signals are collectively known as the Level-1 trigger. If the event passes the Level-1 trigger, it is saved for fuller reconstruction and the *High-Level Trigger* (HLT). Events that pass the HLT are saved for future use and study. By using these triggers, the event rate drops from 20 million events per second to about 50 thousand per second from the Level-1 trigger to approximately 150 events passing the HLT and being saved every second. A certain fraction of events, called minBias events, are collected that have not passed any triggers. This allows for both calibration and searches for new-physics processes that may not pass any traditional trigger.

Even with only 150 events per second that pass the HLT, the raw information collected by each of the detectors inside CMS can be daunting. The next chapter will cover some of the ways in which the raw output from the detectors are collated into physics objects that will be useful in the reconstruction of $t\bar{t}$ events.

Chapter 3

From Detector To Physics

As seen in the previous chapter, the CMS detector is a very powerful device capable of making robust and precise measurements of the particles created in the LHC collisions. These measurements enable tests of the theoretical predictions that have been made about the particles created. However, working with the raw measurements from each subdetector is difficult and time-consuming. It is therefore helpful to combine the different raw measurements into useful *physics objects*. This chapter will cover some of the reconstruction methods and physics objects used in this analysis.

3.1 Particle Flow

The first step to combining the measurements made by the different detectors in CMS is known as *Particle Flow* (PF). Particle flow takes the output from each

detector and matches it to a single particle. For example, a muon is reconstructed by both the silicon tracker and the muon system. Particle flow takes both reconstructed muons, determines they are the same particle, and creates a PF muon containing the information from both detectors. Every physics object has a Particle flow version containing information from multiple detectors.

3.2 Jets

As discussed in Secs. 1.1.2.3 and 2.2.4, a *jet* is the physics object created by quark hadronization. As a jet propagates throughout the detector, the energy spreads out and the reconstructed jet has a roughly conical shape. There are many different algorithms available to reconstruct jets—one class of these algorithms are called k_T-like. These algorithms characterize the distance between two particles (i and j) with a parameter, d_{ij}:

$$d_{ij} = min(k_{T,i}^n, k_{T,j}^n)\frac{\Delta R_{ij}^2}{R^2} \tag{3.1}$$

where $k_{T,i}$ is the transverse momentum of the i-th particle with respect to the beamline, R is a characteristic angular distance and ΔR_{ij} is the distance between particles i and j as in Equation 3.2.

$$\Delta R_{ij} = \sqrt{(\Delta\eta_{ij})^2 + (\Delta\phi_{ij})^2} \tag{3.2}$$

where $\Delta\eta_{ij}$ is the pseudorapdity difference between particles i and j and $\Delta\phi_{ij}$ is the azimuthal[1] difference between the particles. Another important parameter is the the beam distance, d_{iB}:

$$d_{iB} = k_{T,i}^{n} \tag{3.3}$$

The k_T-like algorithms differ in their choice of n. The k_T algorithm has $n = 2$ while the anti-k_T has $n = -2$. Regardless of the choice of n, each algorithm starts with one seed particle, i, and loops through all other particles; if the minimum d_{ij} is less than d_{iB}, the algorithm merges particle j into the jet containing i. The process repeats until d_{ij} is greater than d_{iB}. At that point, the i-th particle is considered a final jet. Eventually all particles are clustered into jets such as in Fig. 3.1.

The anti-k_T algorithm is commonly used in particle physics because it clusters the hardest jets first[2], leading to well resolved jets. However, the anti-k_T algorithm declusters poorly and is therefore not ideal for jet *substructure*.

If n is set to 0 in Eqs. 3.1 and 3.3, the equations simplify to

$$d_{ij} = \frac{\Delta R_{ij}^2}{R^2} \tag{3.4}$$

$$d_{iB} = 1.0 \tag{3.5}$$

[1]The angle in the plane perpendicular to the beamline.

[2]Because $n = -2$, the jets with the largest transverse momenta will be preferentially clustered.

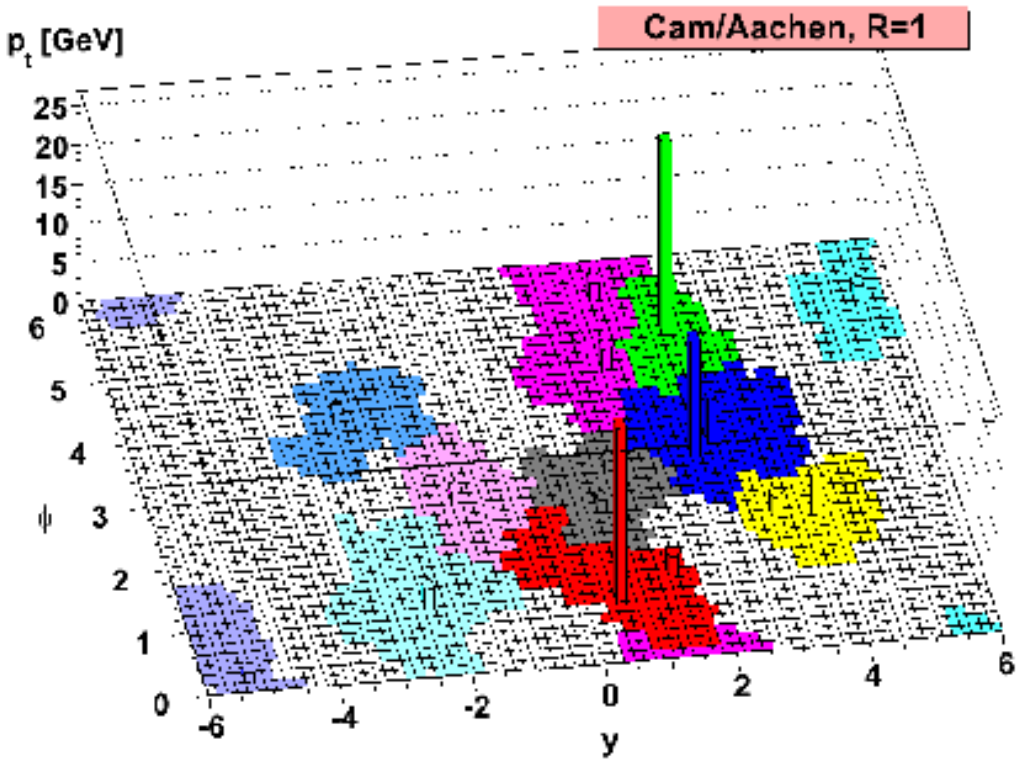

Figure 3.1: Example of the Cambridge/Aachen jet clustering algorithm for R=1 (CA1). Each shaded area is a separate CA jet. Image taken from Ref. [13].

and the resulting algorithm is known as the Cambridge/Aachen (CA) algorithm [19]. The CA algorithm depends only on the spatial characteristics of the jets, so it is ideal for substructure studies, such as reconstructing boosted $t\bar{t}$ jets. Specifically, CA8 jets (jets where R=0.8) were found to work well in this reconstruction.

3.3 Boosted Jets

Figure 3.2 overlays the Feynman diagram of a hadronic top quark decay with a cartoon with the jet cones[3]. As the p_T of the jet increases, the decay becomes more collimated, with the jets from the daughter quarks ("daughter jets") becoming closer to one another. At a certain point[4], the daughter jets of the highly-boosted parent jet come close enough together that they merge. Instead of reconstructing 3 jets, now only one jet is reconstructed, and information from the kinematics of the top decay has been lost. Figure 3.3 shows the top quark decay with one jet that contains the three previous jets. The jet clustering can be undone to recreate these as *subjets* of the merged jet.

3.4 Jet Substructure

The jet clustering process groups different particles into a single jet. If both the detector and the decay process are well understood, the jet's substructure can be used to help reconstruct the decay. The declustering procedure first looks at the two parent subjets[5] for well-separated clusters containing a large fraction of the total jet p_T. If both of the parent subjets pass the declustering criteria, the declustering process is repeated with them; only one of these declusterings is required to succeed.

[3]CA8 jets won't be cones, but a cone-like shape is a decent approximation.
[4]At approximately $p_T = 400$ GeV, a jet becomes known as a *highly-boosted* jet.
[5]The hardest subjets should be clustered last and will be the direct parent of the final jet.

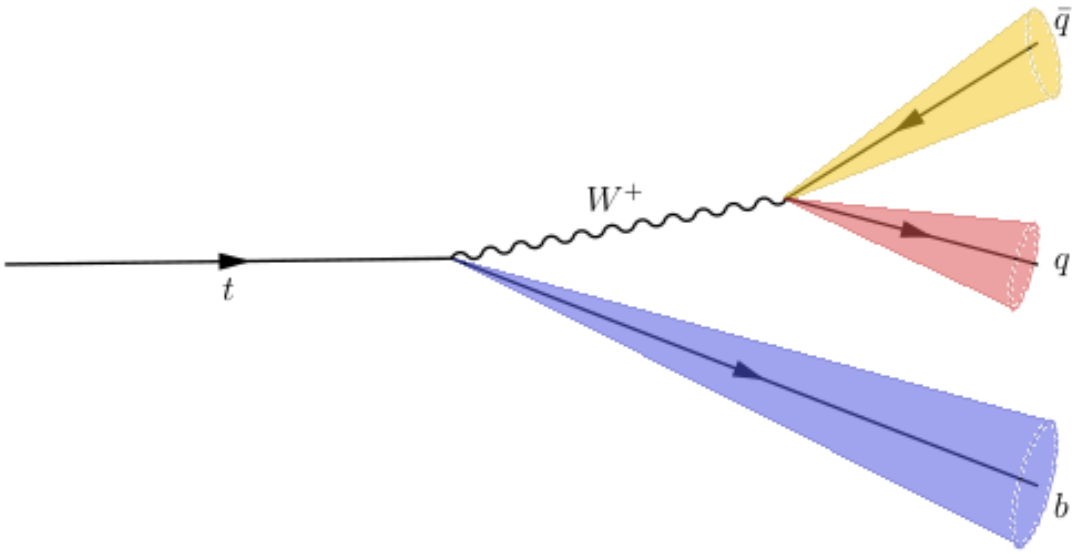

Figure 3.2: Feyman diagram of the hadronic top quark decay with cartoon jet cones overlaid. Three jets are clearly visible: two from the W^+ decay to $q\bar{q}$ and one from the b.

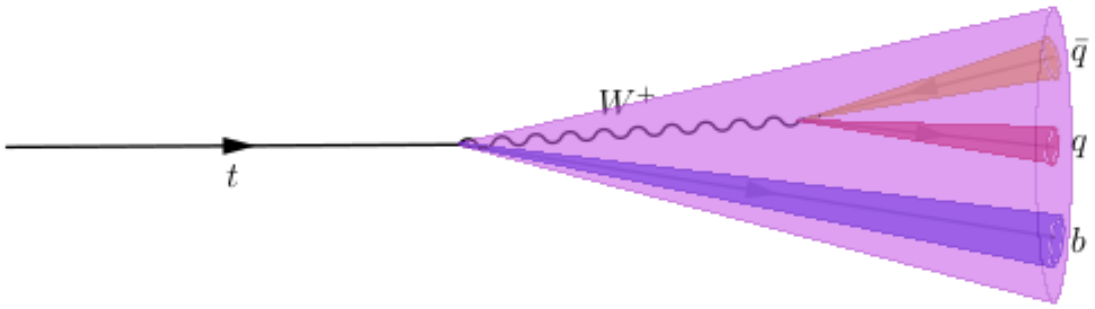

Figure 3.3: Feyman diagram of the boosted hadronic top quark decay with cartoon jet cones overlaid. The three jets are Three jets are clearly visible: two from the W^+ decay to $q\bar{q}$ and one from the b.

This process enables CMS to decompose boosted jets into 3 or 4 subjets, regaining much of the information lost when the jets merged.

3.5 N-Subjettiness

Substructure information can be further leveraged by using a parameter called *N-subjettiness*. The N-subjetiness algorithm defines variables τ_N, where N is the number of subjet axes, as follows:

$$\tau_N = \frac{1}{d_0} \sum_i p_{T_i} \times min(\Delta R_{1,i}, \Delta R_{2,i}, ..., \Delta R_{N,i}) \tag{3.6}$$

where $\Delta R_{j,i}$ is the distance between the subjet axis j and the candidate jet i. Normalizing term d_0, takes p_T into account, with $d_0 = \sum p_{T_i} R_0$, and R_0 is the distance parameter (R) used in the jet clustering algorithm (R = 0.8 for CA8 jets).

N-subjettiness can be used to distinguish between top and non-top jets because hadronic top jets should have three subjet prongs whereas the QCD background will preferentially have one prong. A τ_N value near 0 means the jet likely has N subjets or fewer; a value near 1 means the jet likely has more subjets. The ratio of τ_3/τ_2 (known as τ_{32} and itself bound between 0 and 1) has been shown to provide much better discrimination between top and QCD than either τ_3 or τ_2 and is used in this analysis.

3.6 Subjet btagging

Because of quark confinement, b-quark from the top decay[6] has to form a bound state. The resulting B-hadrons are unstable and decay by the weak force. The nature of the weak force allows the hadron to travel many millimeters, often one or more centimeters, before decaying. When these decays are reconstructed, they do not appear to originate from the interaction point (or primary vertex), but rather at a displaced, secondary vertex. The process by which particles are identified as having come from b-jets is called *btagging*. As with jet clustering, there are many different btagging algorithms, but the Combined Secondary Vertex (CSV) algorithm has been shown to work well with subjets. A perfectly reconstructed boosted ttbar event will have two subjet btags.

The next chapter will detail how each of these physics objects and algorithms is used to selectively reconstruct $t\bar{t}$ jets while suppressing QCD background.

[6]See for example Fig. 3.2.

Chapter 4

Analysis

4.1 Analysis Strategy

The boosted hadronic cross section measurement begins by selecting events with two boosted CA8 jets with $p_T > 400$ GeV from the 2012 run of the LHC. The clustering sequence is rewound by two steps so that each jet is broken down into up to four subjets. The maximum b discriminator value from among all subjets is compared to the CSV medium operating point, and a *b tagged jet*[1] is a CA8 jet which has at least one subjet that passes this requirement.

The jet with the largest p_T is termed the *leading jet* and is required to be consistent with containing 3 subjets through a requirement on τ_{32} [20]. The *subleading jet*, or the jet with the second highest p_T, is required to have a jet mass consistent with a

[1] This is a definition specific to this Thesis.

top jet. The event is then categorized according to how many btags are present.

Data/MC scale factors are calculated and presented for both subjet b tagging and N-subjettiness. These scale factors are measured in a statistically independent μ+jets sample and applied to the hadronic $t\bar{t}$ MC.

A sample of $t\bar{t}$ MC events is fit together with a data-driven Non-Top MultiJet (NTMJ) background to extract the $t\bar{t}$ yield. The fit is a binned maximum likelihood fit of leading jet mass in bins of leading jet p_T.

The raw differential $t\bar{t}$ cross section in bins of leading top jet p_T is presented. Additionally, the resulting distribution is unfolded to parton-level with RooUnfold using Singular Value Decomposition (SVD) to provide the measurement of the true differential $t\bar{t}$ cross section for comparison with theoretical expectations.

4.2 Samples

The analysis was carried out using the full CMS 8 TeV dataset consisting of 19.7 fb^{-1} pp events collected from the LHC in 2012 [21]. Events are required to have passed the `HLT_HT750` trigger, having the scalar sum of the jet energies (H_T) > 750 GeV.

Single Muon samples were used to measure the data/MC scale factors for subjet b tagging and N-subjettiness. The single muon data were collected using the single muon trigger HLT `Mu40_eta2p1` which requires a muon with $p_T > 40$ GeV and $|\eta| <$

2.1. The muons are reconstructed using the Particle Flow algorithm [22].

NTMJ events are accounted for using a data-driven method, but QCD MC is used to check agreement. In addition, MC samples of $t\bar{t}$ events are needed for measuring the ratio of data/MC efficiences for subjet b tagging and N-subjettiness, providing the shape and normalization of $t\bar{t}$ in the fit and constructing a response matrix for unfolding.

$t\bar{t}$ events are modeled by a MC sample that was generated using POWHEG [23] with the CT10 parton distribution function and showered with Pythia6 using the Z2* tune [24]. CMS detector effects are simulated using GEANT 4 [25]. These $t\bar{t}$ events were split into three Monte Carlo samples by $M_{t\bar{t}}$ to increase the number of events in the boosted regime as top quark p_T is directly correlated with $M_{t\bar{t}}$.

4.3 Selection

The event selection is similar to that found in Ref. [26]. Events are required to contain one primary vertex with $|z| < 24$cm and $N_{DOF} > 6$. Pileup must be taken into account and subtracted. This subtraction procedure consists of two components: charged-hadron subtraction (chs) and neutral-hadron subtraction (nhs). Charged hadrons that do not originate from the primary vertex are discarded before particle flow jets (the precursor to CA8 jets) are created. The remaining neutral component of pileup is then removed by applying an area-based correction as described in Ref. [27].

The complete selection after pile-up subtraction requires two CA8 jets with $p_T >$ 400 GeV and $|\eta| < 2.4$. Both jets are also required to have masses consistent with a top quark jet in the range of 140-250 GeV. A $t\bar{t}$ event requires the leading jet $\tau_{32} <$ 0.55. In addition, to guarantee the jet is infra-red safe, the leading jet must have τ_{21} less than 0.1.

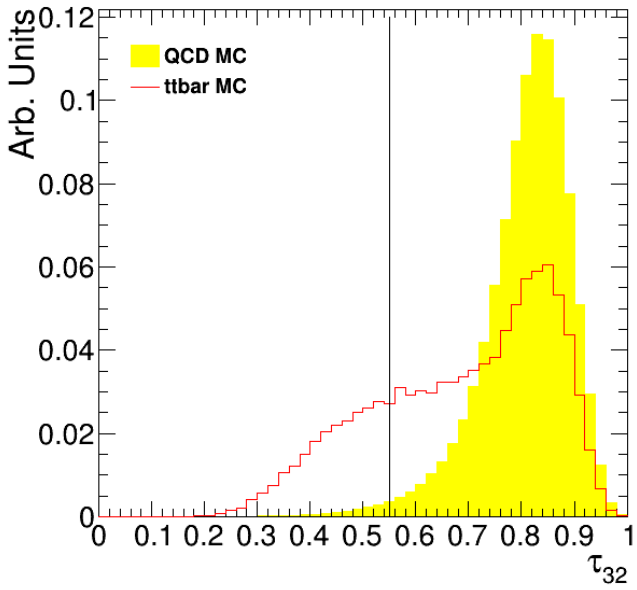

Figure 4.1: Normalized τ_{32} distribution for $t\bar{t}$ MC and QCD MC. The cut at $\tau_{32} =$ 0.55 maximizes $t\bar{t}$ while minimizing NTMJ.

The number of events passing each selection cut as well as the $t\bar{t}$ MC selection efficiency is given in Tab. 4.1.

Events passing this selection are further subdivided according to the number of subjet b tags they have. To define a subjet b tag, we take the declustered subjets and select the maximum btagging discriminant from them and require this value to exceed

Requirement	N_{Data}	ϵ_{MC}
Initial	65576549	—
Trigger	4027708	0.005
Kin. Cuts	4027708	1
$M_1 > 100GeV$	1057544	0.718
$M_2 \in [140, 250]$ GeV	88540	0.372
$Jet_1\tau_{32} < 0.55$	5469	0.398
$Jet_1\tau_{21} > 0.1$	5469	1
$M_1 < 250GeV$	4758	0.979
Final	4758	0.0015

Table 4.1: Event counts and MC selection efficiency for analysis-level cuts.

the Combined Secondary Vertex (CSV) medium operating point of 0.679. Thus the number of subjet b tags actually counts the number of CA8 jets with at least one subjet b tag.

Events are divided into three exclusive categories: events with no jets containing a subjet passing the btagging requirement, events with 1 jet (leading or subleading) containing subjets passing the btagging requirement, and events with 2 or more jets containing subjets passing the btagging requirement

These are referred to as 0 btags, 1 btag, and 2 btags, respectively. Table 4.2 gives the events counts after selection for these three categories.

	0 btags	1 btag	2 btags
$t\bar{t}$	92.9± 4.9	288.2± 8.4	204.1±7.0
NTMJ	3181.1±18.1	831.0± 9.3	77.2±2.8
Total	3274.1±18.8	1119.1±12.5	281.3±7.6
Data	3085	1103	265

Table 4.2: Event counts and statistical uncertainties for data, $t\bar{t}$, and NTMJ background, for $p_T > 400$ GeV prior to fitting.

Finally, to get the differential $t\bar{t}$ cross section we divide our sample into the following 5 regions based on the leading jet p_T and independently measure the $t\bar{t}$ yield in each region: 400 GeV $\leq p_T <$ 500 GeV, 500 GeV $\leq p_T <$ 600 GeV, 600 GeV $\leq p_T <$ 700 GeV, 700 GeV $\leq p_T <$ 800 GeV, and 800 GeV $\leq p_T <$ 1200 GeV.

Yields for these regions are given in Table 4.3.

400	0 btags	1 btag	2 btags
$t\bar{t}$	28.4±2.7	92.1±4.8	66.9±3.9
NTMJ	820.0±9.1	213.9±4.6	21.5±1.5
Total	848.5±9.5	306.0±6.7	88.4±4.2
Data	738	282	83

500	0 btags	1 btag	2 btags
$t\bar{t}$	34.0±3.0	107.1±5.1	81.9±4.6
NTMJ	1254.7±11.7	328.9±6.0	30.2±1.9
Total	1288.7±12.0	435.9±7.9	112.1±4.9
Data	1064	425	114

600	0 btags	1 btag	2 btags
$t\bar{t}$	18.0±2.1	56.2±3.7	37.8±3.0
NTMJ	554.4±6.8	145.4±3.5	14.5±1.1
Total	572.3±7.1	201.7±5.1	52.3±3.2
Data	668	210	33

700	0 btags	1 btag	2 btags
$t\bar{t}$	6.8±1.5	23.0±2.3	11.9±1.8
NTMJ	321.9±5.9	85.7±3.0	6.1±0.8
Total	328.7±6.0	108.7±3.8	18.0±1.9
Data	360	103	22

800	0 btags	1 btag	2 btags
$t\bar{t}$	5.7±1.2	9.8±1.6	5.6±1.2
NTMJ	230.2±5.4	57.1±2.7	4.8±0.8
Total	235.9±5.6	66.8±3.1	10.4±1.5
Data	255	83	13

Table 4.3: Event counts and statistical uncertainties for data, $t\bar{t}$, and background for 400 GeV$< p_T <$ 500 GeV (top left), 500 GeV$< p_T <$ 600 GeV (top right), 600 GeV$< p_T <$ 700 GeV (middle left), 700 GeV$< p_T <$ 800 GeV (middle right), 800 GeV$< p_T <$ 1200 GeV (bottom); prior to fitting.

4.3.1 Jet Parameters

Jet parameters for both the leading and subleading jet will be shown in the following figures for both the *tagged region* ($\tau_{32} < 0.55$) and the *antitagged region* ($\tau_{32} > 0.55$) . As shown in Fig. 4.10, τ_{32} for the leading jet in the tagged region has to be treated specially because NTMJ is defined by requiring $\tau_{32} > 0.55$. In order to plot the τ_{32} parameter, we took the normalization of the region where $\tau_{32} > 0.55$ and scaled the region with $\tau_{32} < 0.55$ to match this value. This was only done in order to visualize the τ_{32} distribution—the shape of this parameter does not matter in this analysis.

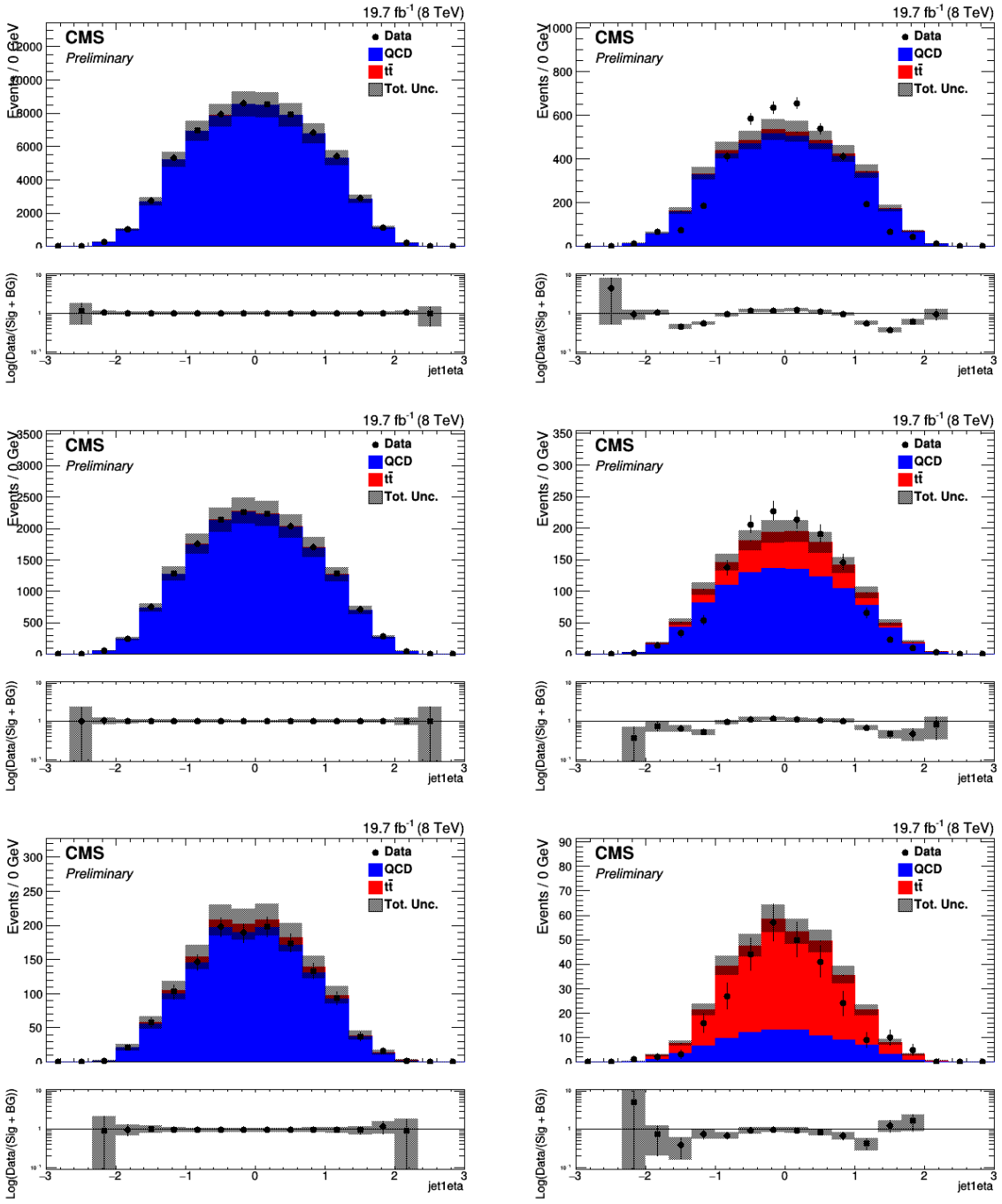

Figure 4.2: Pseudorapidity distribution of the leading jets in the antitag (left) and the tag (right) region for 0 btags (top), 1 btag (middle) and 2 btags (bottom) after selection, but before fitting. The discrepancy of the leading jet η is well understood and comes from the τ_{32} requirement.

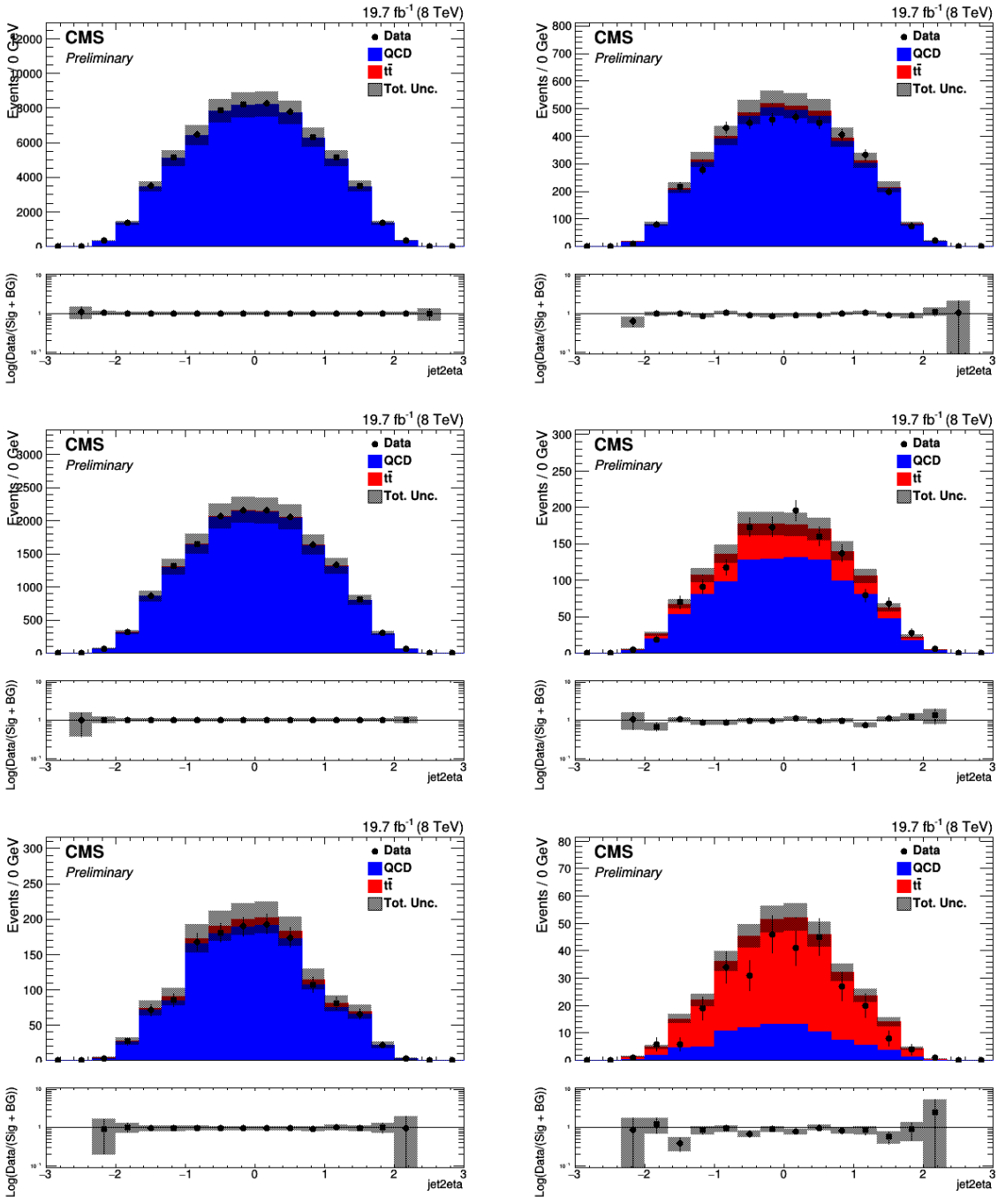

Figure 4.3: Pseudorapidity distribution of the subleading jets in the antitag (left) and the tag (right) region for 0 btags (top), 1 btag (middle) and 2 btags (bottom) after selection, but before fitting.

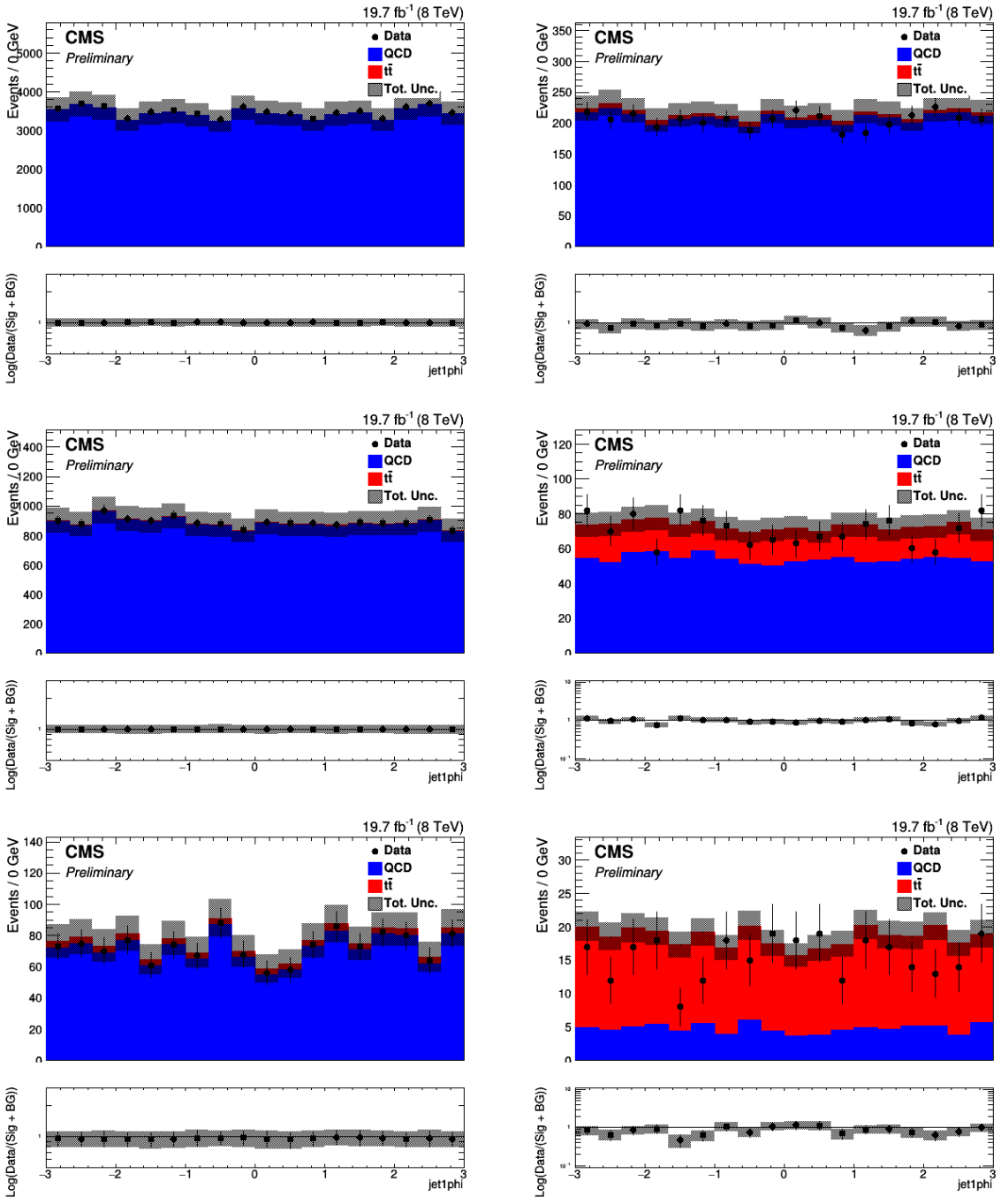

Figure 4.4: Phi distribution of the leading jets in the antitag (left) and the tag (right) region for 0 btags (top), 1 btag (middle) and 2 btags (bottom) after selection, but before fitting.

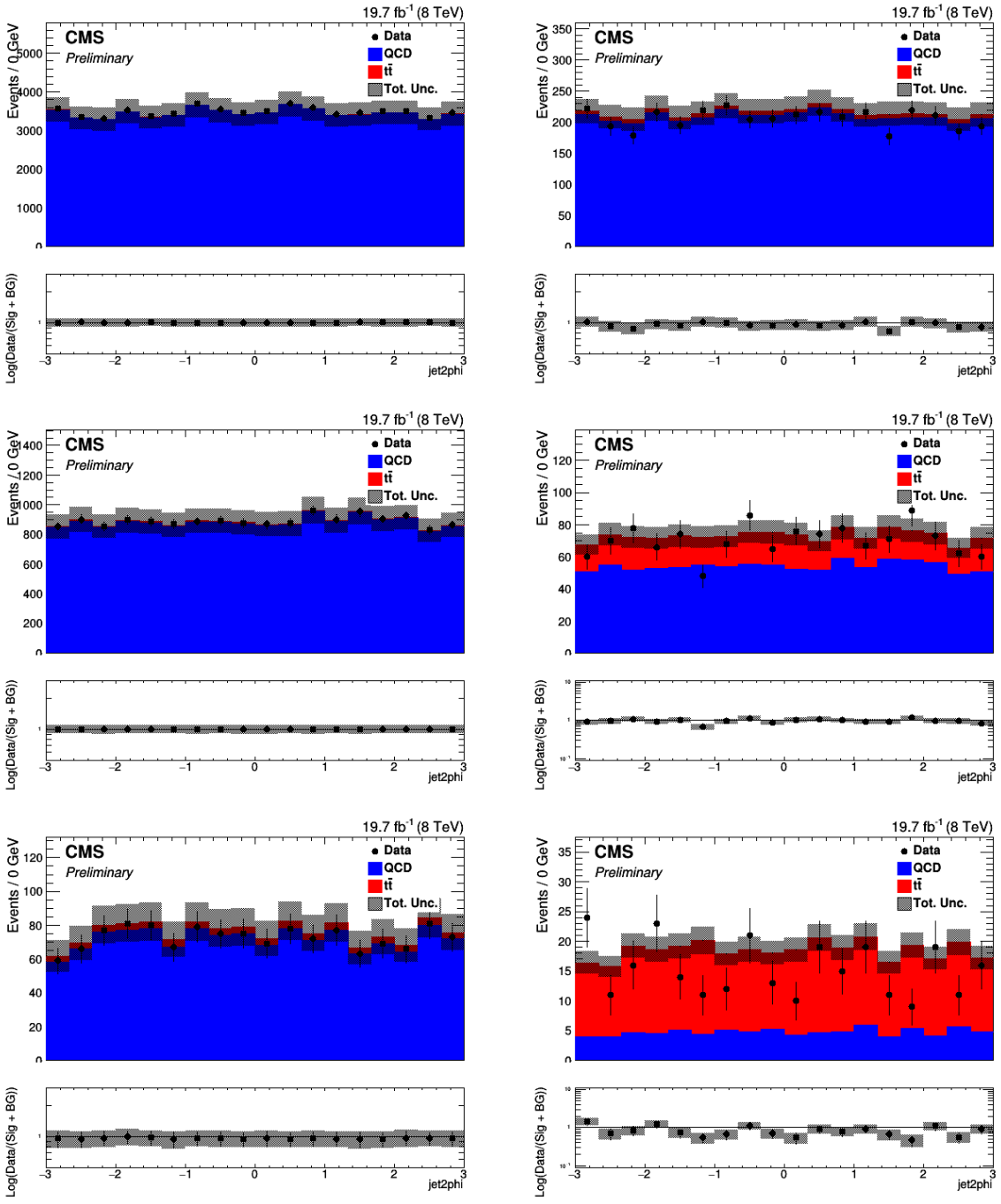

Figure 4.5: Phi distribution of the subleading jets in the antitag (left) and the tag (right) region for 0 btags (top), 1 btag (middle) and 2 btags (bottom) after selection, but before fitting.

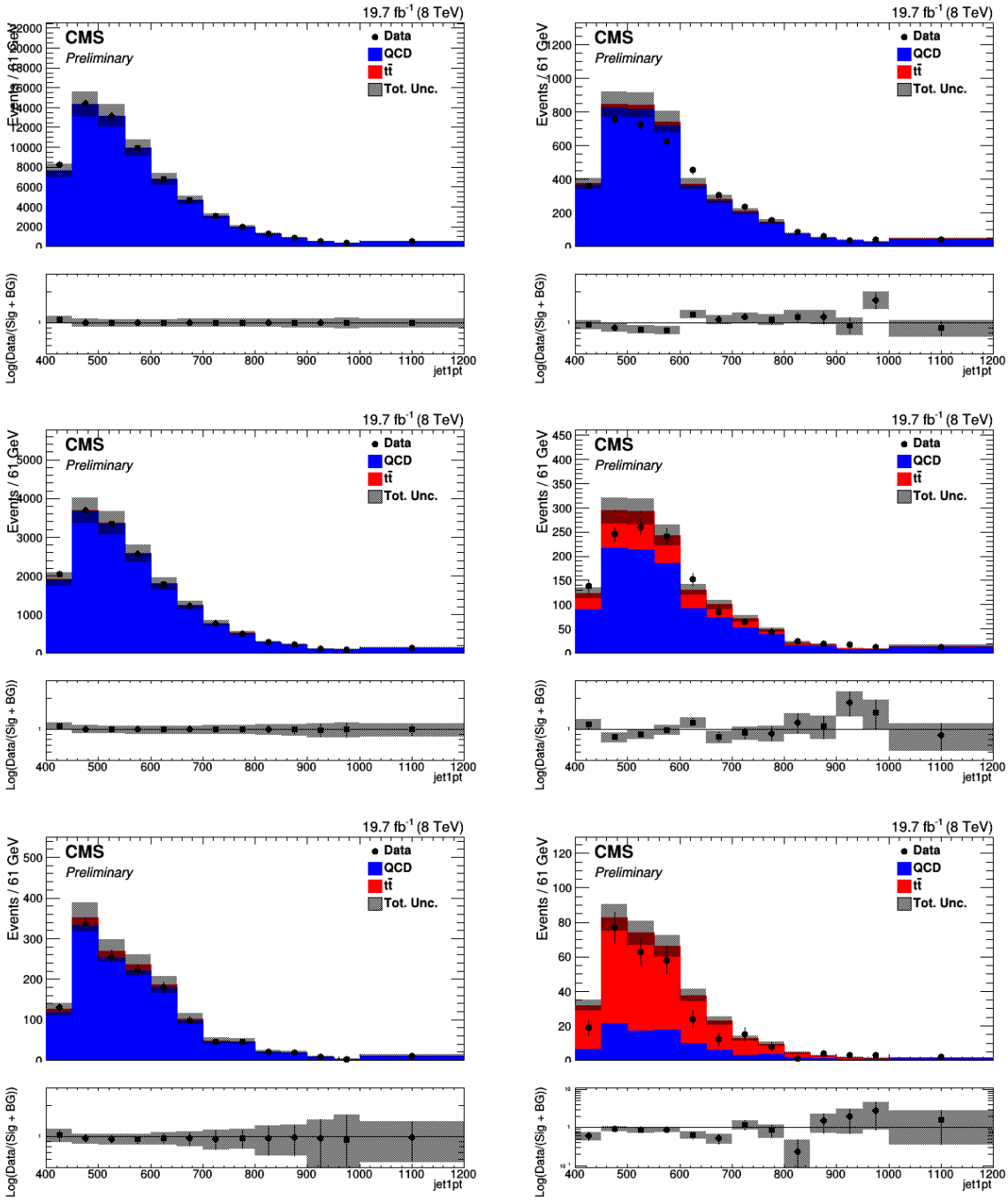

Figure 4.6: Transverse momentum distribution of the leading jets in the antitag (left) and the tag (right) region for 0 btags (top), 1 btag (middle) and 2 btags (bottom) after selection, but before fitting.

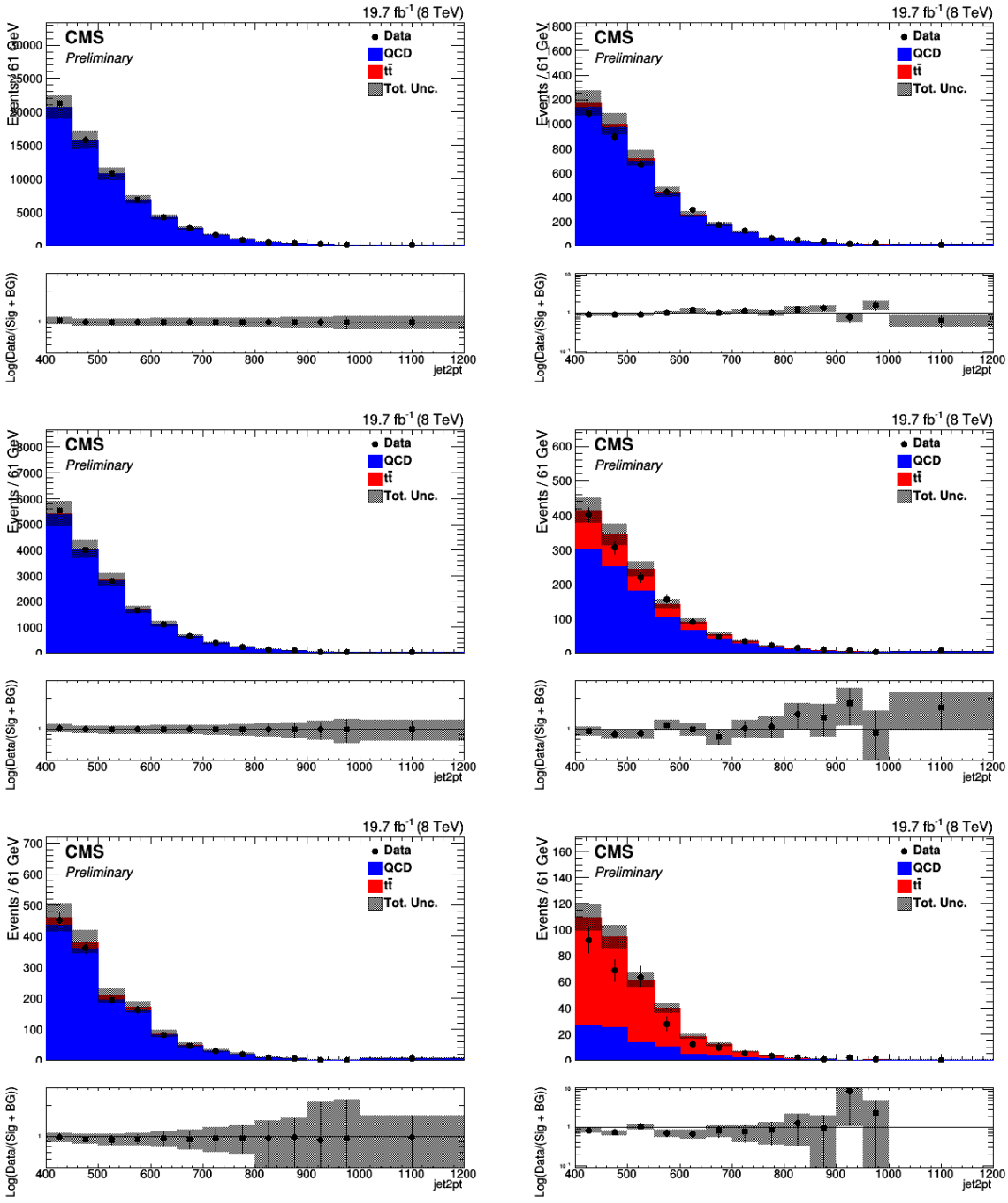

Figure 4.7: Transverse momentum distribution of the subleading jets in the antitag (left) and the tag (right) region for 0 btags (top), 1 btag (middle) and 2 btags (bottom) after selection, but before fitting.

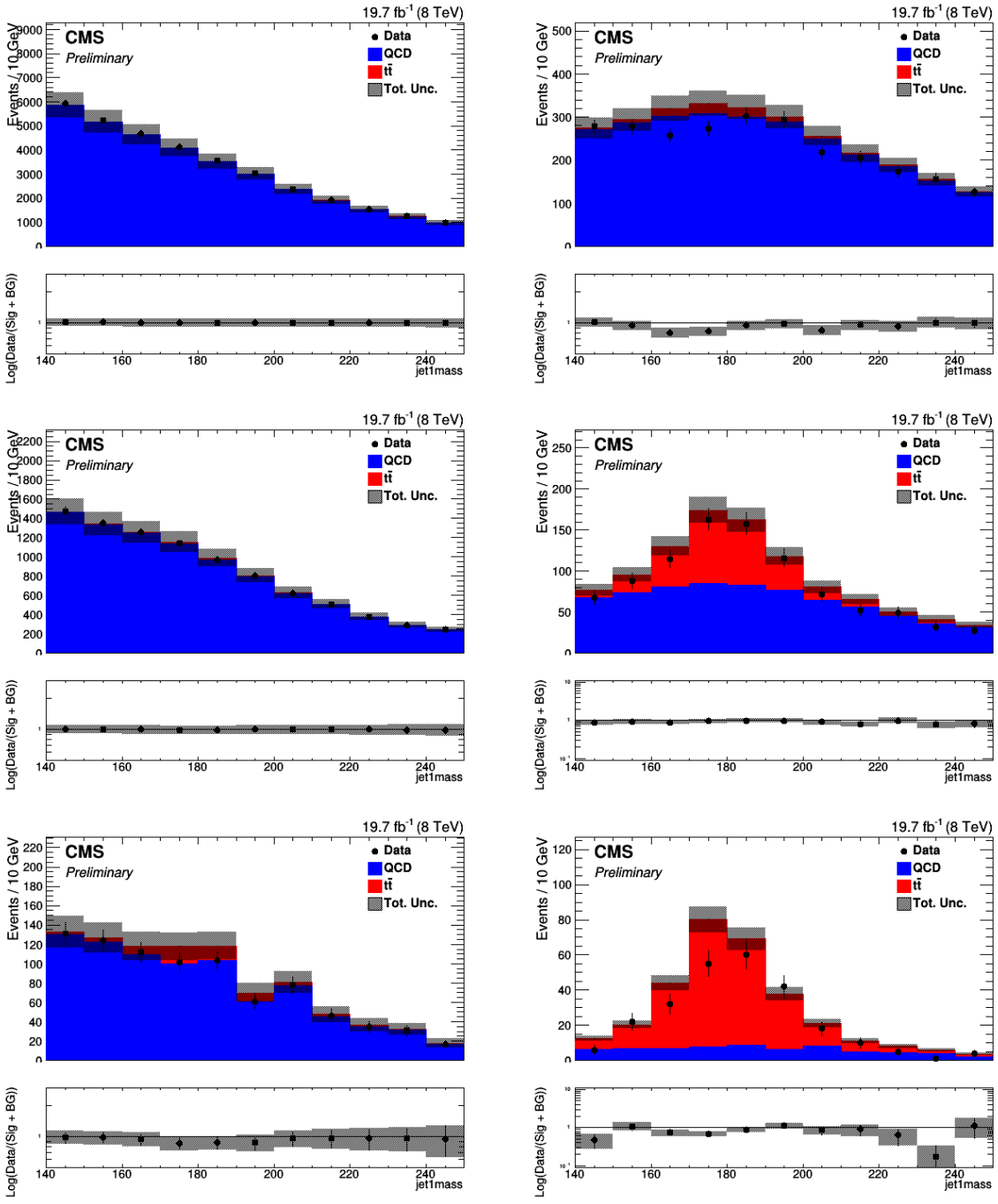

Figure 4.8: Mass distribution of the leading jets in the antitag (left) and the tag (right) region for 0 btags (top), 1 btag (middle) and 2 btags (bottom) after selection, but before fitting.

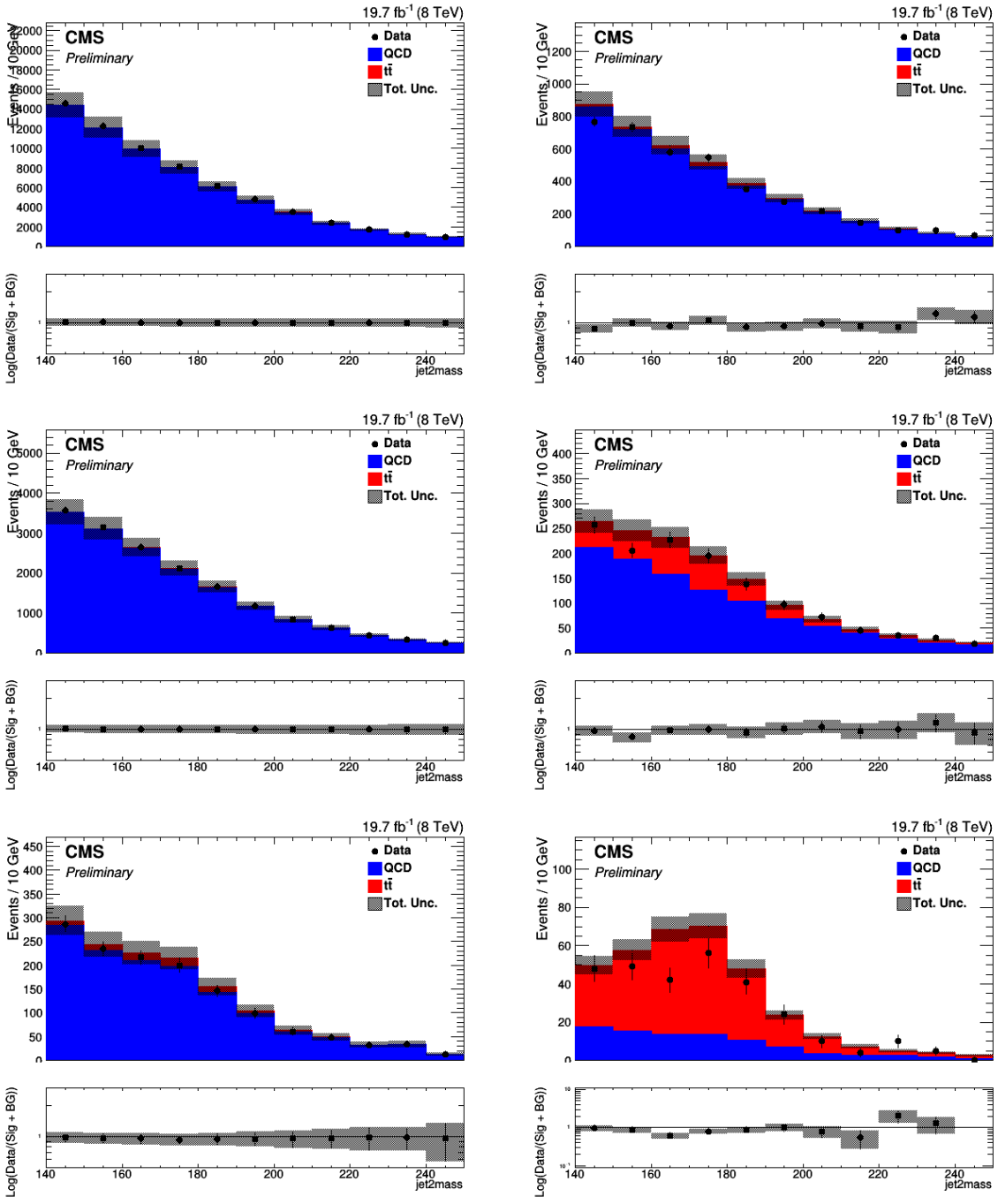

Figure 4.9: Mass distribution of the subleading jets in the antitag (left) and the tag (right) region for 0 btags (top), 1 btag (middle) and 2 btags (bottom) after selection, but before fitting.

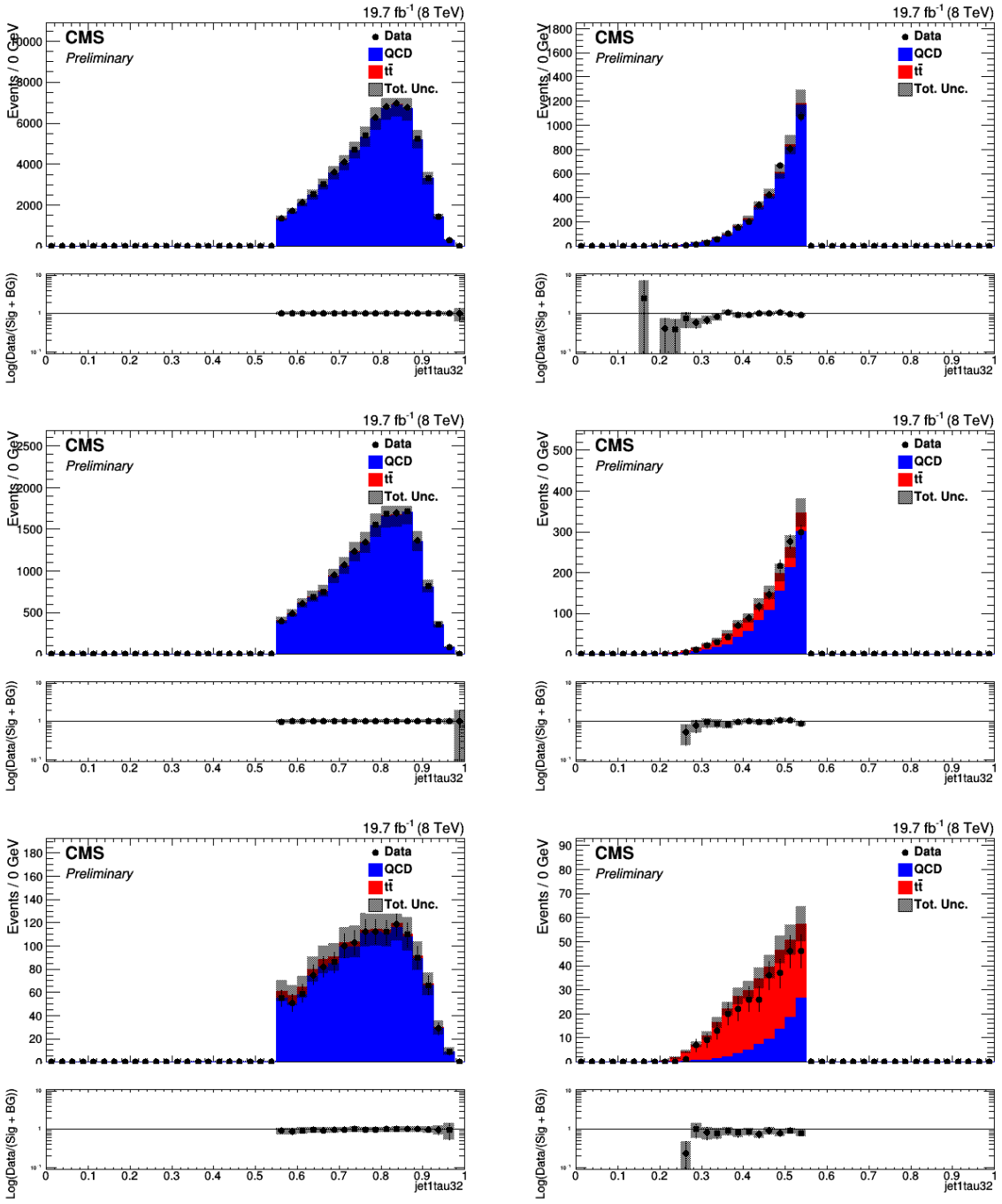

Figure 4.10: τ_{32} distribution of the leading jets in the antitag (left) and the tag (right) region for 0 btags (top), 1 btag (middle) and 2 btags (bottom) after selection, but before fitting. For the antitag distribution, the normalization for NTMJ is taken from the sidebands.

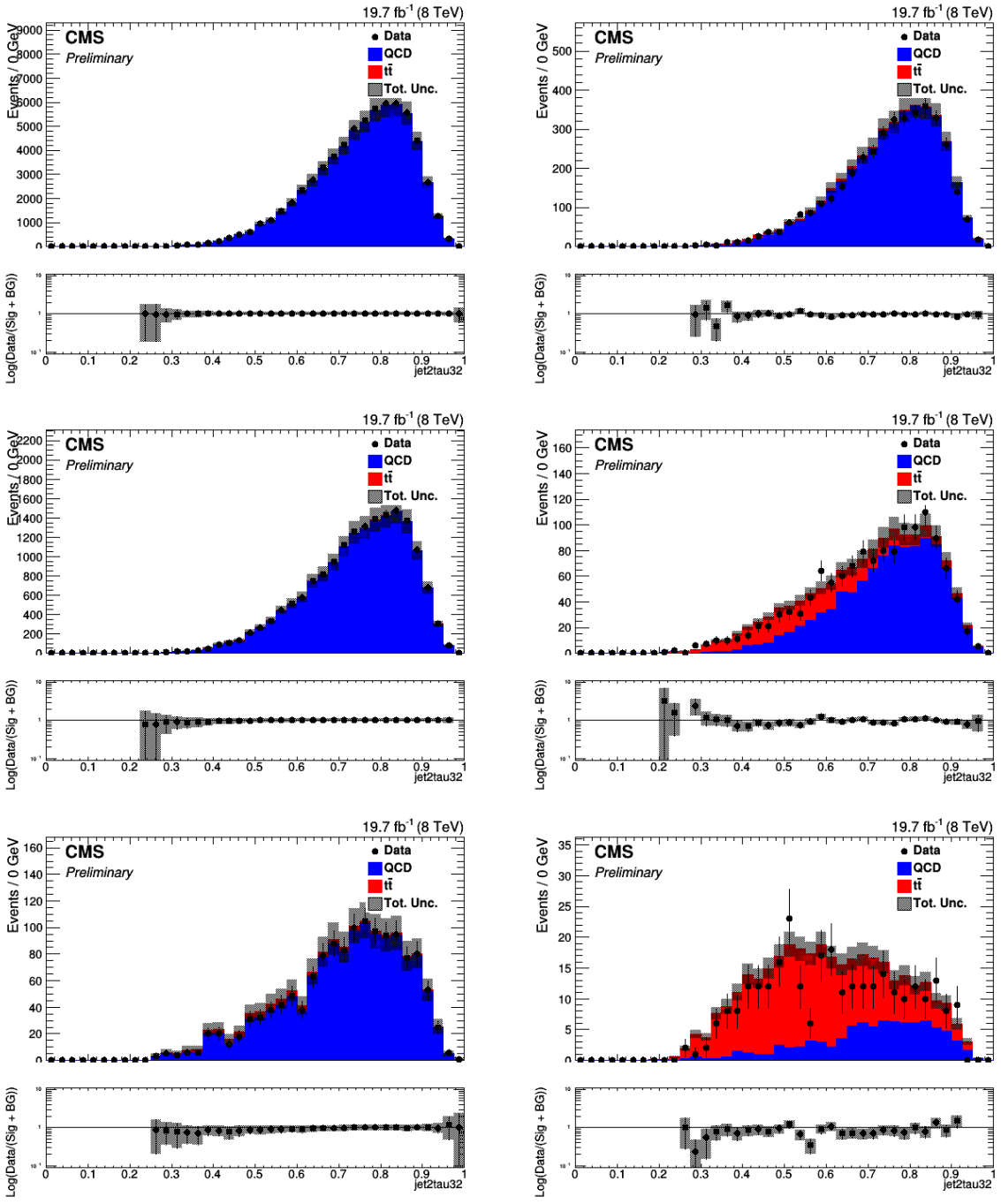

Figure 4.11: τ_{32} distribution of the subleading jets in the antitag (left) and the tag (right) region for 0 btags (top), 1 btag (middle) and 2 btags (bottom) after selection, but before fitting.

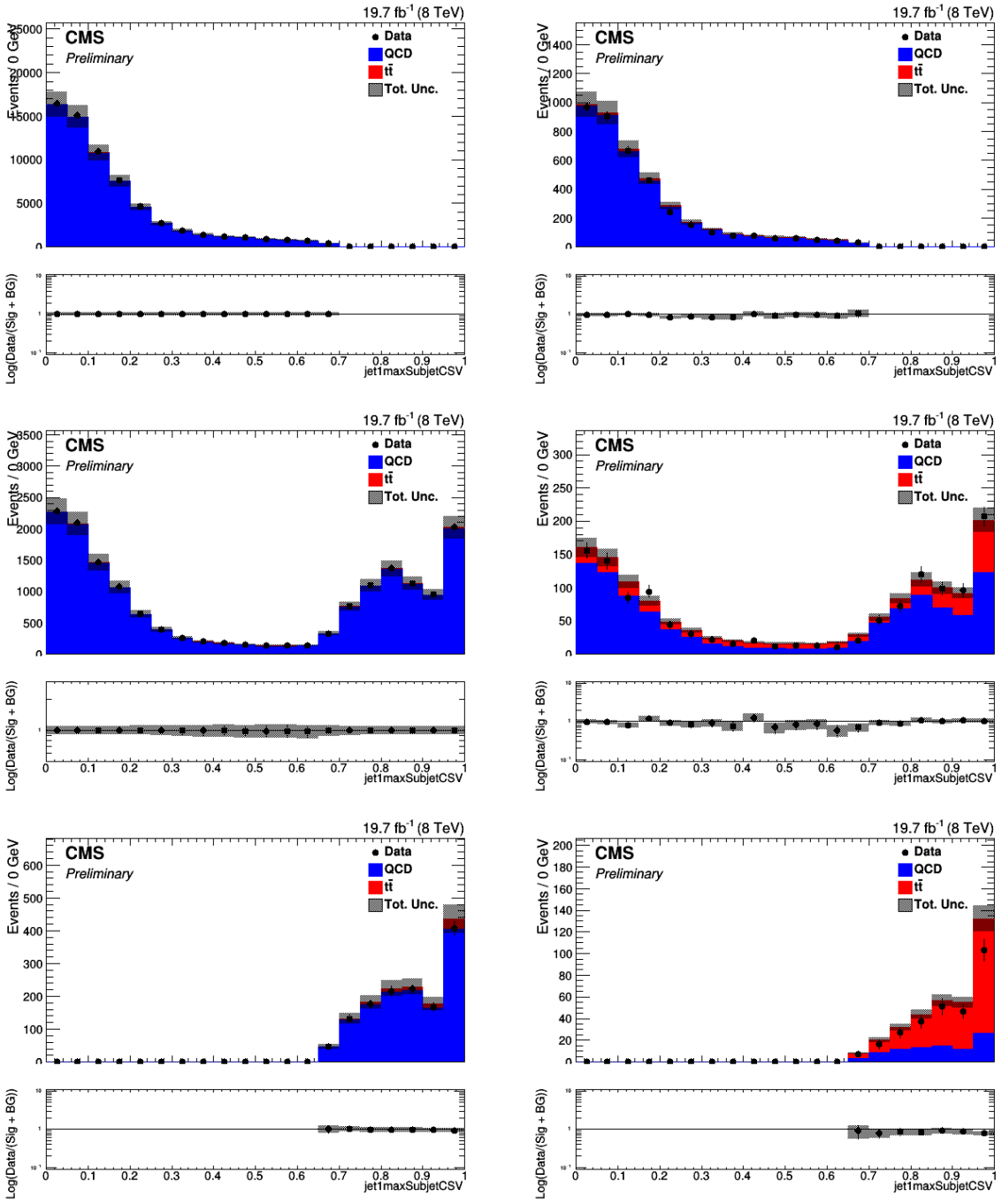

Figure 4.12: Maximum CSV from all subjets of the leading jets in the antitag (left) and the tag (right) region for 0 btags (top), 1 btag (middle) and 2 btags (bottom) after selection, but before fitting.

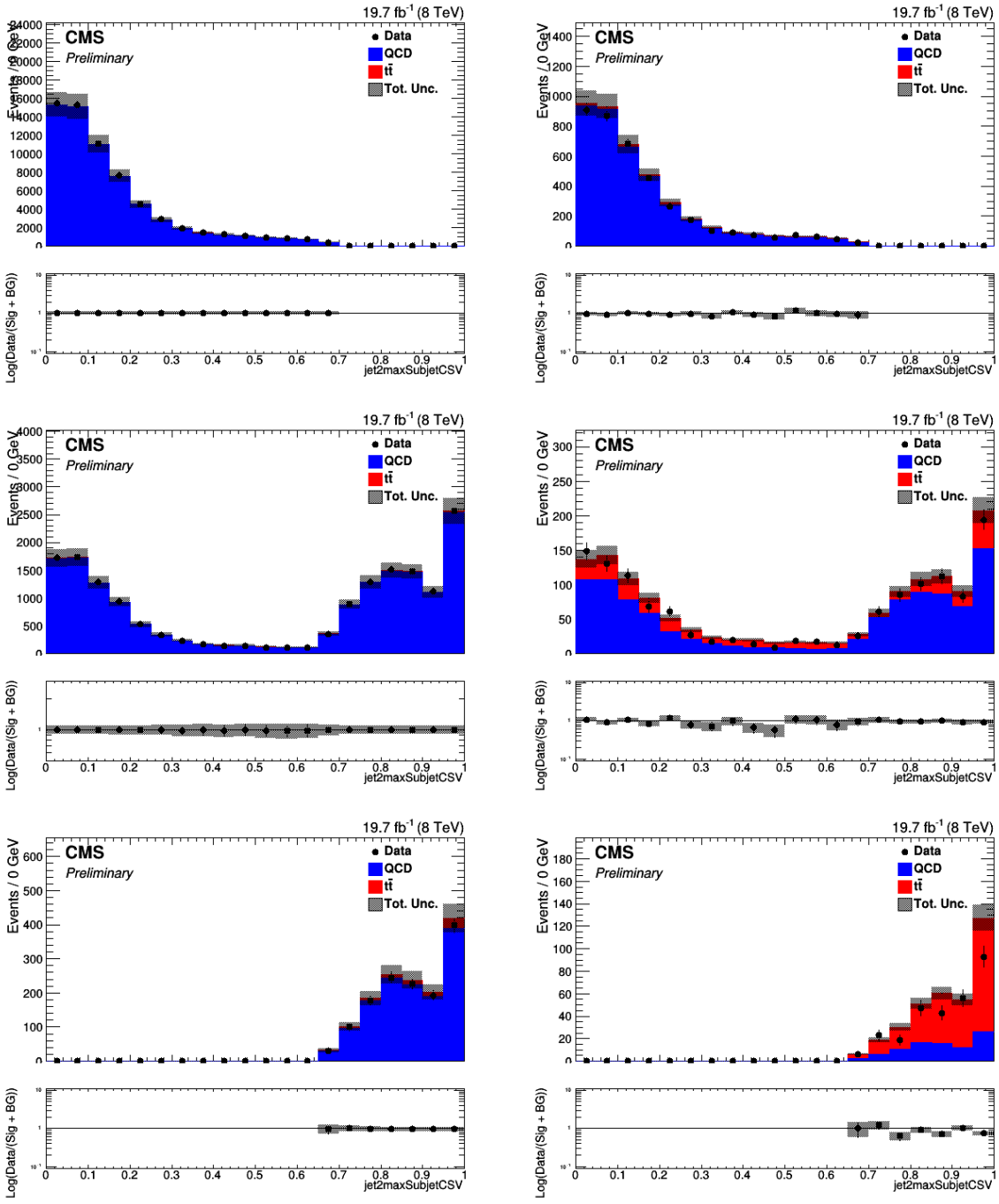

Figure 4.13: Maximum CSV from all subjets of the subleading jets in the antitag (left) and the tag (right) region for 0 btags (top), 1 btag (middle) and 2 btags (bottom) after selection, but before fitting.

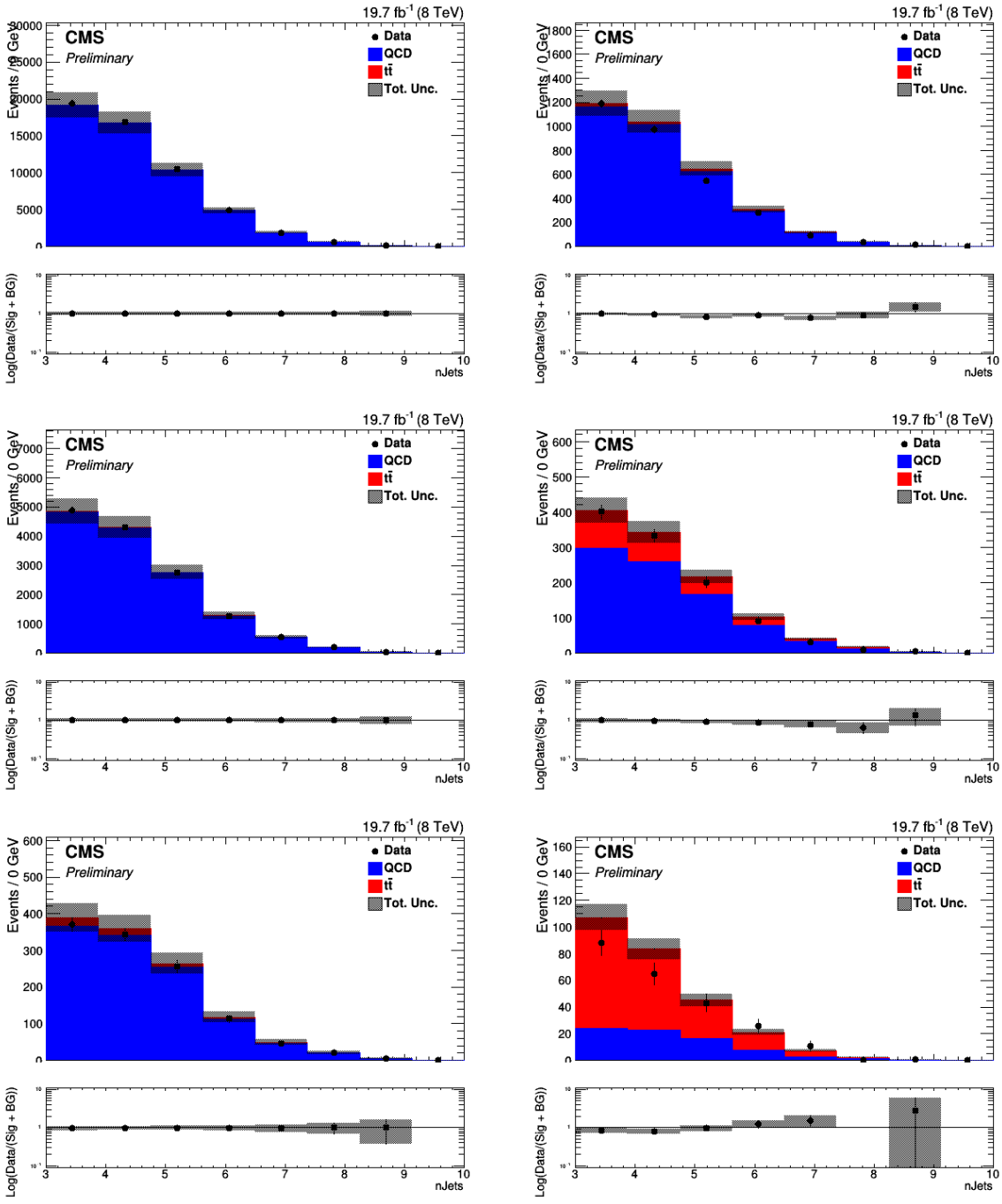

Figure 4.14: Number of jets spectrum in the antitag (left) and the tag (right) region for 0 btags (top), 1 btag (middle) and 2 btags (bottom) after selection, but before fitting.

4.4 Scale Factor Measurement

Due to a combination of mismodeled hadronic shower in PYTHIA (needed to describe some b tagging and substructure variables) and mismodeled detector effects in full simulation, the efficiency of a cut in data may not be the same as in MC simulation. We correct this by applying a *scale factor* (SF) to MC events to make them match data. A statistically independent μ+jets sample is used to calculate the efficiency of a cut in both data and MC. The ratio of data efficiency to MC efficiency is the scale factor to be applied to signal $t\bar{t}$ MC in all-jet events as seen in Eq. 4.1. As important is the *uncertainty* on the scale factor because it determines the degree of understanding of the modeling of the signal. Of particular interest to this analysis are scale factors for subjet btagging and N-subjettiness.

$$\text{SF} = \frac{\epsilon^{Data}}{\epsilon^{MC}} \tag{4.1}$$

The muon+jets dataset is used to derive these two scale factors as it is statistically independent from the hadronic selection used in the rest of this analysis. The following preselection is applied to all muon+jets events, requiring an insolated muon identified by the Particle Flow algorithm with $p_T > 45$ GeV. The isolation requirement is either $\Delta R > 0.5$ between the muon and the closest jet or a relative $p_T > 25$GeV between the muon and the closest jet (2D isolation). The hadronic jet also requires $\tau_{21} > 0.1$ for IR safety, and the leptonic "top jet" must be in the top mass window [140,250]

GeV to kinematically match the hadronic analysis.

The muon is identified by the Particle Flow algorithm and then combined with missing E_T to create a W candidate. This W candidate is then combined with the jet with the second highest p_T to make a leptonic top candidate. (The highest p_T jet is the fully merged hadronic top.)

The data and Monte Carlo events are then further divided into the following four disjunct regions: bt— Hadronic Jet $\tau_{32} < 0.55$ and at least one subjet b tag, b— Hadronic Jet $\tau_{32} > 0.55$ and at least one subjet b tag, t— Hadronic Jet $\tau_{32} < 0.55$ and no subjet b tags, and f— Hadronic Jet $\tau_{32} > 0.55$ and no subjet b tags.

These regions can be seen in Fig. 4.15.

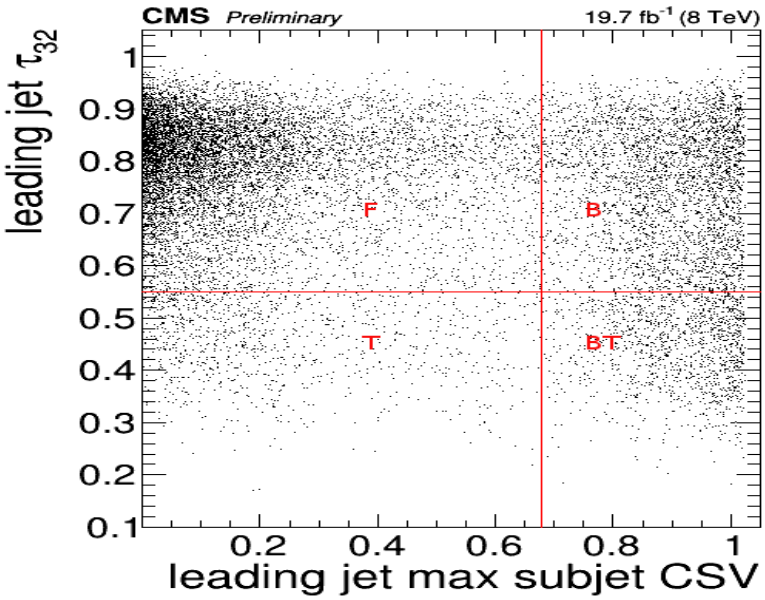

Figure 4.15: Hadronic Jet Subjet CSV Score vs Hadronic Jet N-subjettiness.

The leading jet mass distribution of each of the four regions in Fig. 4.15 is fit

simultaneously with two components using RooFit. The signal top jet is represented by a Gaussian peaking at 170 GeV that is common to all regions; there is also a broad Gaussian background in each region that behaves independently of the background in any other regions. The normalizations of the top signal pdfs are given by the following set of equations:

$$N_{bt} = N_{total} \times \epsilon_b \times \epsilon_t \tag{4.2}$$

$$N_b = N_{total} \times \epsilon_b \times 1 - \epsilon_t \tag{4.3}$$

$$N_t = N_{total} \times 1 - \epsilon_b \times \epsilon_t \tag{4.4}$$

$$N_f = N_{total} \times 1 - \epsilon_b \times 1 - \epsilon_t \tag{4.5}$$

where N_{total} is the total number of signal events and N_{bt}, N_b, N_t, and N_f are the number of events in a given region. ϵ_b is the reconstruction efficiency for a hadronic jet with a btagged subjet and ϵ_t is the reconstruction efficiency for a hadronic jet passing the N-subjettiness requirement. The results of these fits to MC simulation can be found in Fig. 4.16. Results for data are shown in Fig. 4.17. The data/MC scale factors are derived as the ratio ϵ^{data} / ϵ^{MC} and are given in Table 4.4 along with the errors calculated from those provided by the fit.

Single Top MC samples were originally included but have been shown to have a negligible effect; they are not included in the analysis.

	$\tau_{32} < 0.55$	subjet b tag
Data eff.	0.667 ± 0.061	0.604 ± 0.045
MC eff.	0.577 ± 0.034	0.793 ± 0.024
Data/MC	0.841 ± 0.051	1.048 ± 0.070

Table 4.4: Data/MC scale factors for analysis-level cuts.

4.5 Background Estimate

The primary background to hadronic $t\bar{t}$ production is Non-Top MultiJet (NTMJ) production. The primary component of NTMJ events is QCD, and it dominates $t\bar{t}$ production, so having an accurate measure of the NTMJ background is vitally important to any $t\bar{t}$ cross section measurement in the all-jets channel. There are many ways to approach determining the amount of NTMJ background, but one common problem is that many parameters good at separating NTMJ from $t\bar{t}$, such as τ_{32}, are correlated with parameters of interest such as the invariant mass of the leading jet used in the fit to extract the $t\bar{t}$ yield. Both methods using Minimum-Pairwise Mass (MPM) and Large-R Isolation and Separation Adjustment (LISA) were attempted but rejected because of this correlation[2].

For this reason, we use a modified ABCD method, referred to as the Alphabet method, to estimate background in the presence of this correlation between τ_{32} and leading jet mass (Fig. 4.18).

The background estimate is made by running on the hadronic Run I data and requiring events to pass the following selection: leading jet $\tau_{21} > 0.1$ for IR safety,

[2]LISA is not a real method. MPM is however and was rejected.

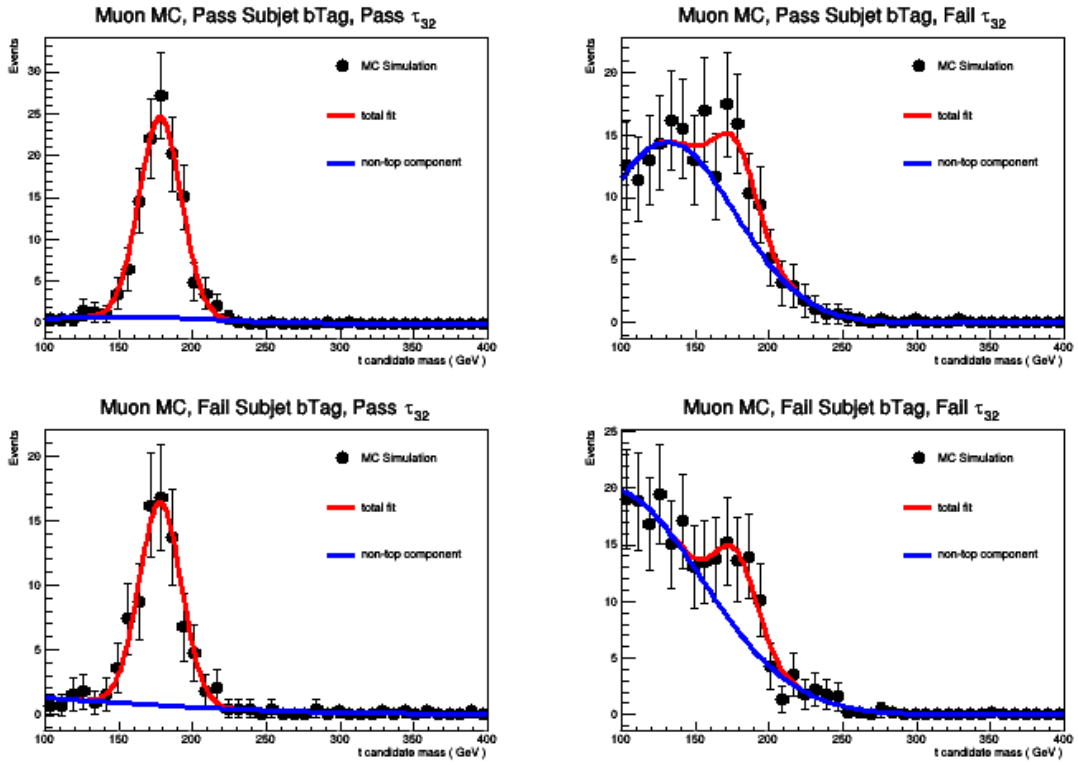

Figure 4.16: Signal and background fit to $t\bar{t}$ Monte Carlo events to determine the MC scalefactors for subjet btagging and N-subjettiness. (Top left) bt events (Top right) b events (Bottom left) t events (Bottom right) f events.

leading jet mass in the window [100,400] GeV (This is the variable used in the fit, so we broaden the mass cut to include the "sidebands" of the top peak, and subleading jet mass in the top window [140,250] GeV.

This selection is also applied to $t\bar{t}$ MC, and events that pass are subtracted from the sample to be fit to give a NTMJ-enriched sample.

The τ_{32}-vs-mass phase space of this $t\bar{t}$-depleted data is broken up into 6 regions. These are labeled *A*, *B*, *C*, *D*, *E*, and *F* and visualized in Fig. 4.18. Regions *A*, *B*, *D*, and *F* are the sidebands in leading jet mass. Regions *A* and *B* have leading jet

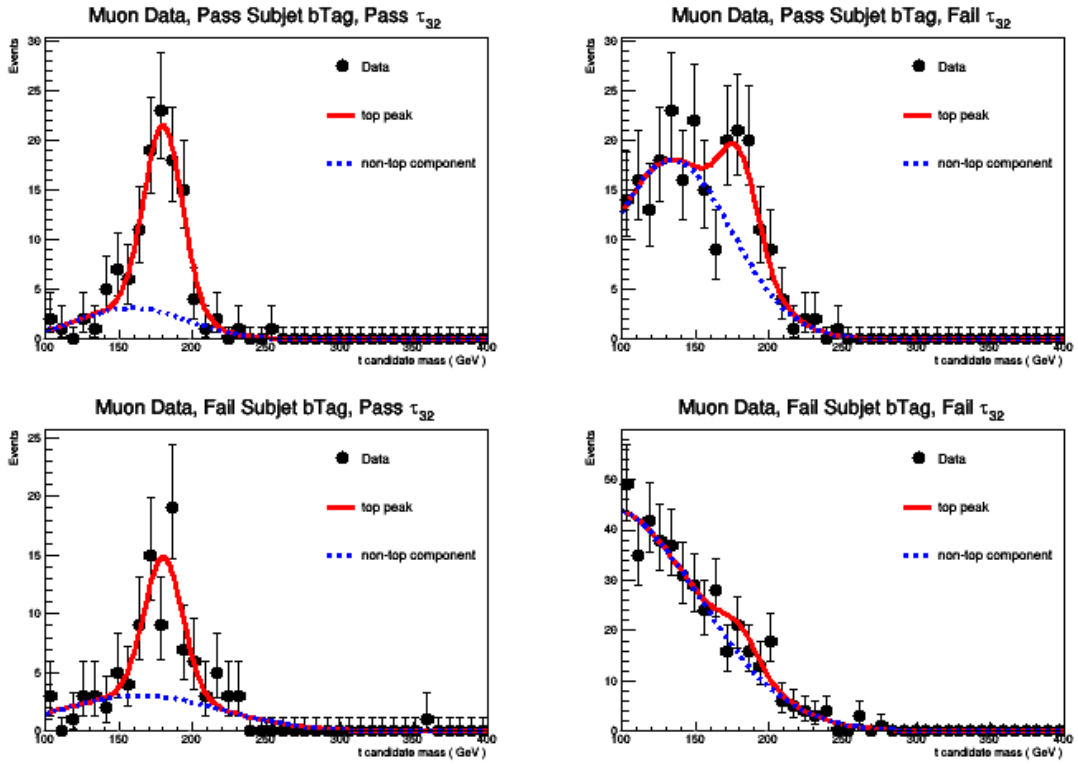

Figure 4.17: Signal and background fit to $t\bar{t}$ semi-muonic data to determine the MC scalefactors for subjet btagging and N-subjettiness. (Top left) bt events (Top right) b events (Bottom left) t events(Bottom right) f events.

mass in the window [100,140] GeV while regions E and F have leading jet mass in the window [250,400] GeV. The signal region (regions C and D) is where the leading jet mass is consistent with a top jet ([140,250]) GeV.

Regions B, D, and F where events pass the $\tau_{32} < 0.55$ requirement are considered top-like, while regions with events failing that cut (regions A, C, and E) are considered to be NTMJ.

The ratio of events passing the τ_{32} selection to events failing that selection is measured for the sidebands. That is, the number of events in region B is divided by

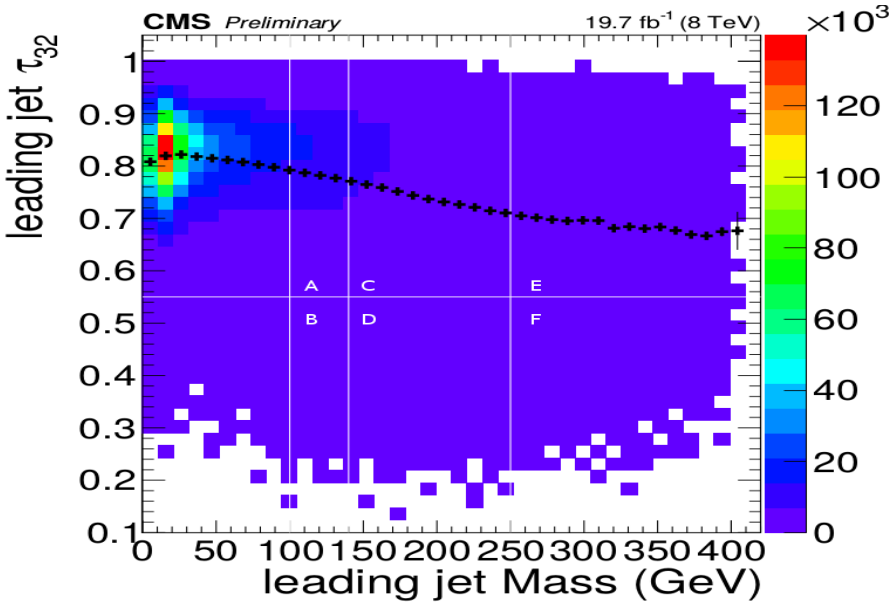

Figure 4.18: Plot of τ_{32} vs Leading Jet Mass. The black points show the profile distribution of τ_{32} in bins of jet mass showing the correlation between τ_{32} and jet mass. The plot is subdivided into the regions used in the Alphabet method. Note that regions B, D, and F are in the low-τ_{32} tail of the τ_{32} distribution of each vertical slice of this plot. For this reason, the Alphabet method can approximate the correlation between "pass" and "fail" events (here below and above the line, repectively) with a smooth linear or parabolic function.

the number of events in region A and the number of events in region F is divided by the number of events in region E. This allows an estimate of top-like events (or events that would pass the top selection) given a sample of NTMJ events (those that fail the τ_{32} cut). This ratio is measured in the sidebands and a quadratic function, the *transfer function*, is used to interpolate between low and high jet mass to get an estimate of the pass/fail ratio in the signal region. These are then applied to events failing the τ_{32} cut to estimate the NTMJ in the signal region. This procedure is done for each of the p_T regions considered in the analysis. This defines a transfer fuction

parametrized as a function of leading jet p_T and leading jet mass. In the differential $t\bar{t}$ cross section measurement, every event that fails the τ_{32} cut is multiplied by an event weight from the transfer function, given its jet mass. The sum of resulting weighted events constitute the data-driven NTMJ background.

This same weighting is applied to $t\bar{t}$ Monte Carlo events to account for mistagged signal events that fail the τ_{32} cut and thus are in the anti-tagged region *C*. These events need to be removed from the anti-tag region in order not to overestimate the amount of NTMJ in the signal region *D*.

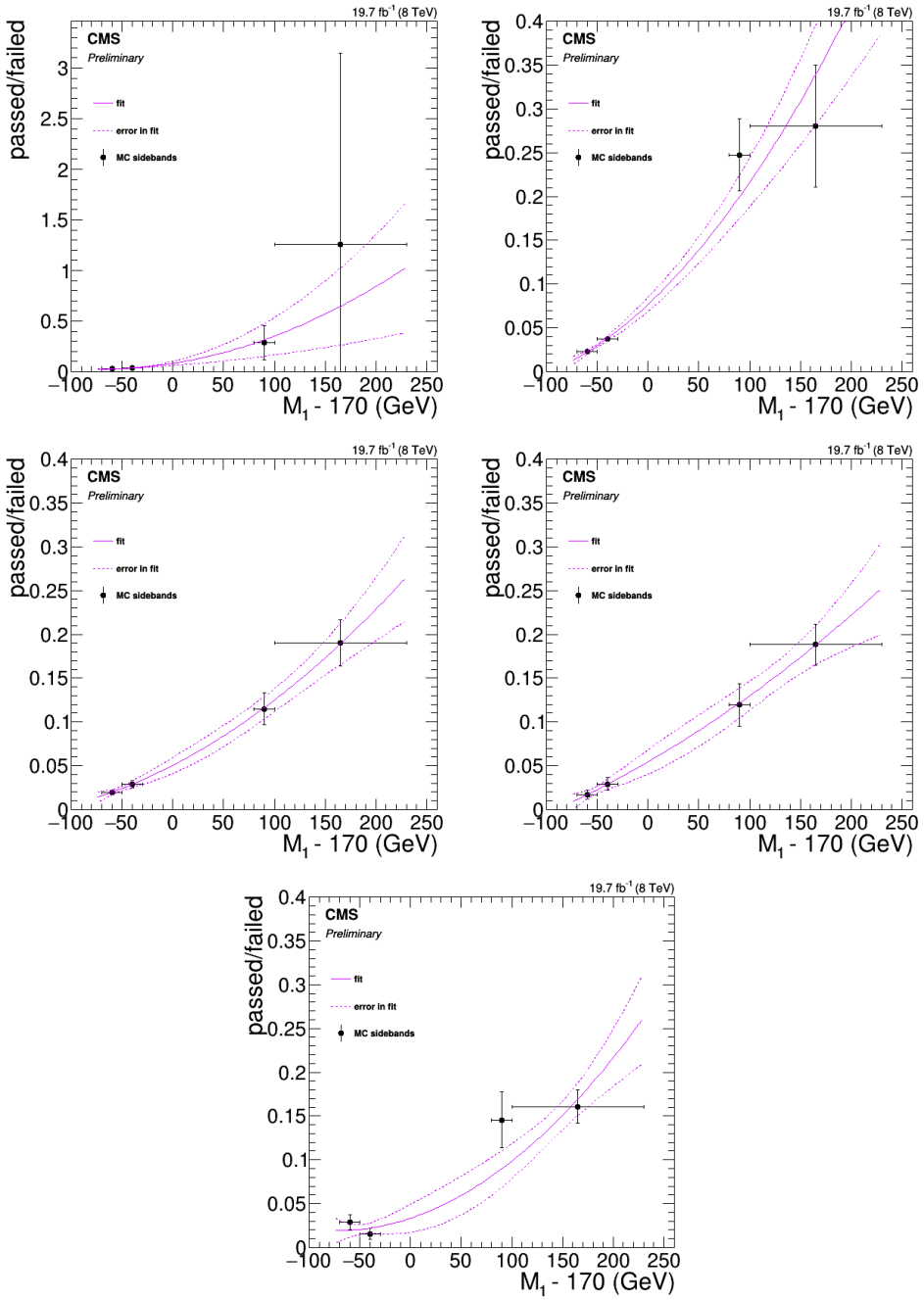

Figure 4.19: Quadratic Transfer Functions for leading jet p_T between 400 and 500 GeV (top left), 500 and 600 GeV (top right), 600 and 700 GeV (middle left), 700 and 800 GeV (middle right), and 800 and 1200 GeV (bottom).

4.6 Systematics

4.6.1 Jet Energy Corrections (JES & JER)

In CMS, the calorimeter response to particles is not linear and therefore corrections must be applied to jets to allow the proper mapping of the measured jet energy deposition to the particle-level jet energy. These jet energy corrections are comprised of Jet Energy Scale (JES) and Jet Energy Resolution (JER) corrections. For both of these corrections, the approach recommended by the Jet and Missing Energy (JME) group in Ref. [28] is followed. Corrections were make for anti-k_T 5 and 7 (AK5 and AK7, respectively) jets, but not for CA8 jets. It has been previously found in Ref. [29] that the AK7 corrections can be used for CA8 jets. Jets are corrected using the `START53_V27` set of corrections for MC samples and the `Winter14_V5_DATA` set of corrections for data. The JES uncertainty is estimated by varying the jet energy in bins of jet p_T and jet η using the AK7 corrections and adding additional 3% flat uncertainty to the result. Because jets in simulation have a better JER than jets in data, the jet energy is smeared in simulation by 10%. The JER uncertainty is evaulated by smearing an additional η-dependent amount as determined by CMS. Templates with JES shifted $\pm 1\sigma$ are provided to the fitter, as are templates with JER shifted $\pm 1\sigma$. An additional flat 1% uncertainty is provided to the fit to account for using AK7 corrections for CA8 jets.

4.6.2 Pileup (pu)

The effect of pileup has been discussed previously in section 4.3. The number of vertices in each event is compared between data and MC simulation. The simulation is then weighted with a function describing this difference. To estimate the pileup uncertainty, the min bias cross section is shifted up and down by 5% and the pileup procedure rerun. The result is a pair of templates corresponding to $\pm 1\sigma$ pileup uncertainty.

4.6.3 PDF uncertainty (pdf)

The $t\bar{t}$ MC samples were generated using the CT10 PDF sets. The value of the PDF directly relates to the cross section generated, so the decision was made to externalize these from the fit. The CT10 PDF set uncertainty was found by following the PDF4LHC prescription given in Ref. [30]. The fit was repeated with the nominal template replaced by either the $\pm 1\sigma$ PDF template and rerun. The difference between the fit result using the nominal template and the fit result using the PDF-shifted template was taken as the PDF uncertainty.

4.6.4 Subjet btagging and N-subjettiness Scale Factors (btag & nsub)

The amount of $t\bar{t}$ reconstructed is dependent on both the subjet btagging and N-subjettiness efficiencies discussed and calculated in Sec. 4.4. These were found to vary between data and MC simulation, resulting in the scale factors shown in Table 4.4. Each of the scale factors is independently shifted $\pm 1\sigma$ and saved as a set of templates that are provided to the fitter.

4.6.5 Transfer Function (xfer)

The procedure by which transfer functions were created was described in Sec. 4.5. Each p_T region considered in the differential measurement is provided with a pair of templates representing a $\pm 1\sigma$ shift in the transfer function for that region. These are allowed to change independently of each other.

4.6.6 Q^2 (q2)

The uncertainty on the modeling of the hard-scatter process ($Q^2 = m_{top} + \sigma p_T^2$), was accounted for with dedicated $t\bar{t}$ samples where the renormalization and factorization scales are varied up or down by twice their nominal values. The fit was rerun with the $t\bar{t}$ MC replaced by the Q^2-shifted MC. The difference between the fit with the nominal $t\bar{t}$ MC and the Q^2-shifted $t\bar{t}$ MC is taken as the uncertainty due to the

Q^2 systematic.

4.6.7 Luminosity (lum)

A flat uncertainty of 2.6% is applied to the $t\bar{t}$ MC to account for luminosity uncertainty [21].

4.6.8 Trigger (trig)

Uncertainty in the `HLT_HT750` trigger was also considered, but found to be negligible.

4.6.9 Systematic Strength

The relative strength of each of these systematics for the inclusive $p_T > 400$ GeV sample is shown in Tab. 4.5. For each of the 3 btag subsets (0, 1, or 2 subjet btags), the strength of a 1 σ deviation in each systematic is shown. This serves to give the relative strength of each systematic only; the fit is not constrained to a 1 σ change.

4.7 Maximum Likelihood Fit

The $t\bar{t}$ yield is extracted from a fit of the invariant mass distribution of the leading jet mass with the sum of $t\bar{t}$ and NTMJ components. The leading jet mass is binned

0 btags	$t\bar{t}$	QCD
JER	0.22	0.00
JES	13.16	0.19
xfer	0.08	18.71
nsub	8.89	0.29
btag	18.59	0.22

1 btag	$t\bar{t}$	QCD
JER	0.33	0.07
JES	13.11	2.75
xfer	0.89	20.88
nsub	7.92	4.69
btag	5.98	0.81

2 btags	$t\bar{t}$	QCD
JER	0.14	0.84
JES	12.69	28.19
xfer	0.62	27.04
nsub	8.19	25.48
btag	10.14	14.79

Table 4.5: Table showing the relative strength of the systematics in the fit for 0 subjet btags (left), 1 subject btag (center), and 2 subjet btags (right). Each entry corresponds to the percent increase/decrease in events for either $t\bar{t}$ or QCD for a 1 σ change in each systematic.

in both p_{T} and number of subjet btags (0,1,2). In each p_{T} region, the btag bins are simultaneously fit using a binned maximum likelihood fit. For each region, the τ_{32}-tagged events are fit simultaneously with antitag events. The binned likelihood for each p_{T} region is given by Equation 4.6.

$$L = \prod_{i}^{btags} \prod_{m}^{mass} \frac{N[i,m]^{n_i} e^{-N[i,m]}}{n_i!} \prod_{j}^{syst} e^{P_j(i,m)} \prod_{k}^{MCstat} e^{P_k(i,m)} \tag{4.6}$$

where n_i is the number of events having i btags, N[i,m] is the number of events in jet-mass bin m of a template with i btags given by Equation 4.7, and $\prod_j^{syst} e^{P_j(i,j)}$ contains the gaussian penalty terms for all systematics. Uncertainties due to limited simulated statistics are included through $\prod_k^{MCstat} e^{P_k(i,m)}$ term: a set of uncorrelated nuisance parameters which allow the expectation in each bin of the background templates to vary within Gaussian constraints.

$$N[i,m] = N^{tag}_{t\bar{t}} \times P^{tag}_{t\bar{t}}(i,m) \; + \; N^{tag}_{NTMJ} \times P^{tag}_{NTMJ}(i,m) \; + \; N^{at}_{t\bar{t}} \times P^{at}_{t\bar{t}}(i,m) \; + \; N^{at}_{NTMJ} \times P^{at}_{NTMJ}(i,m) \quad (4.7)$$

where $N_{t\bar{t}}$ is the number of $t\bar{t}$ events, N_{NTMJ} is the number of NTMJ events and $P_{t\bar{t}}$, and P_{NTMJ} are the pdfs of the $t\bar{t}$ and NTMJ distributions, respectively. Separate distributions are used for events which have passed the N-subjettiness requirement (tag) and those that have failed it (at or antitag). The shape and normalization for $P_{t\bar{t}}$ is taken from the simulated $t\bar{t}$ template, whereas P_{NTMJ} is created by taking events which fail the τ_{32} requirement and weighting them by the transfer function.

Hadronic data and $t\bar{t}$ MC events are required to pass the full selection whereas NTMJ events must pass the full event selection except the τ_{32} cut. All templates are subdivided according to how many subjet b tags they have.

In addition to these nominal templates, a pair of $\pm 1\sigma$ shifted templates are provided for each source of systematic error described in Sec. 4.6. Combine uses these shifted templates to allow the systematic to vary quadratically between $\pm 1\sigma$; outside this range, Combine performs a linear extrapolation. In addition, each bin in the $t\bar{t}$ MC templates as well as the data-driven NTMJ background templates is allowed to vary up and down within its statistical errors, as described by Barlow and Beeston [31]. These are provided to the fit as a set of templates, where each Barlow-

Beeston template matches the nominal template except one bin is shifted by $\pm 1\sigma_{stat}$.

The fit is firstly perfomed inclusively in the leading jet p_T and used to derive the cross section in the fiducial region $p_T > 400$ GeV. The postfit tagged and anti-tagged leading jet mass and p_T are presented in Figs. 4.20 and 4.21 correspondingly. Five sets of probability distribution functions are then used for the differential measurement, with each one representing one of the p_T regions listed in Section 4. The results of the individual fits of these regions (using 1 and 2 subjet btags only) is reported in Figs. 4.22 and 4.23.Additional fits can be found in the Appendix (Sec. 6.1.2).

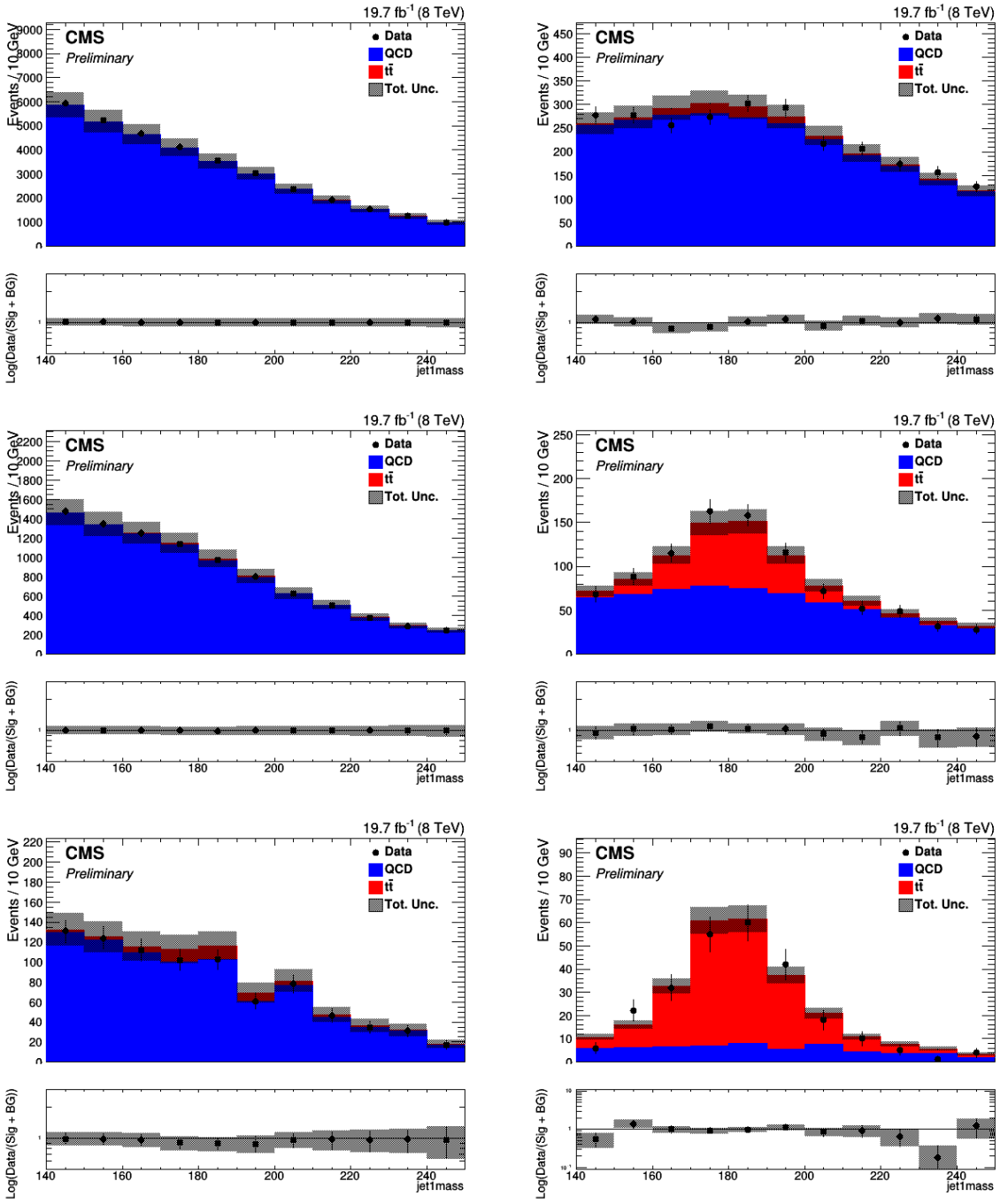

Figure 4.20: (Top Panels) Mass distribution of leading jet in the antitag (left) and tag (right) region for events with 0 btags (top), 1 btag (middle) and 2 btags (bottom) for the inclusive ($p_T > 400$ GeV) sample. (Bottom Panels) Ratio of data to combined signal + background.

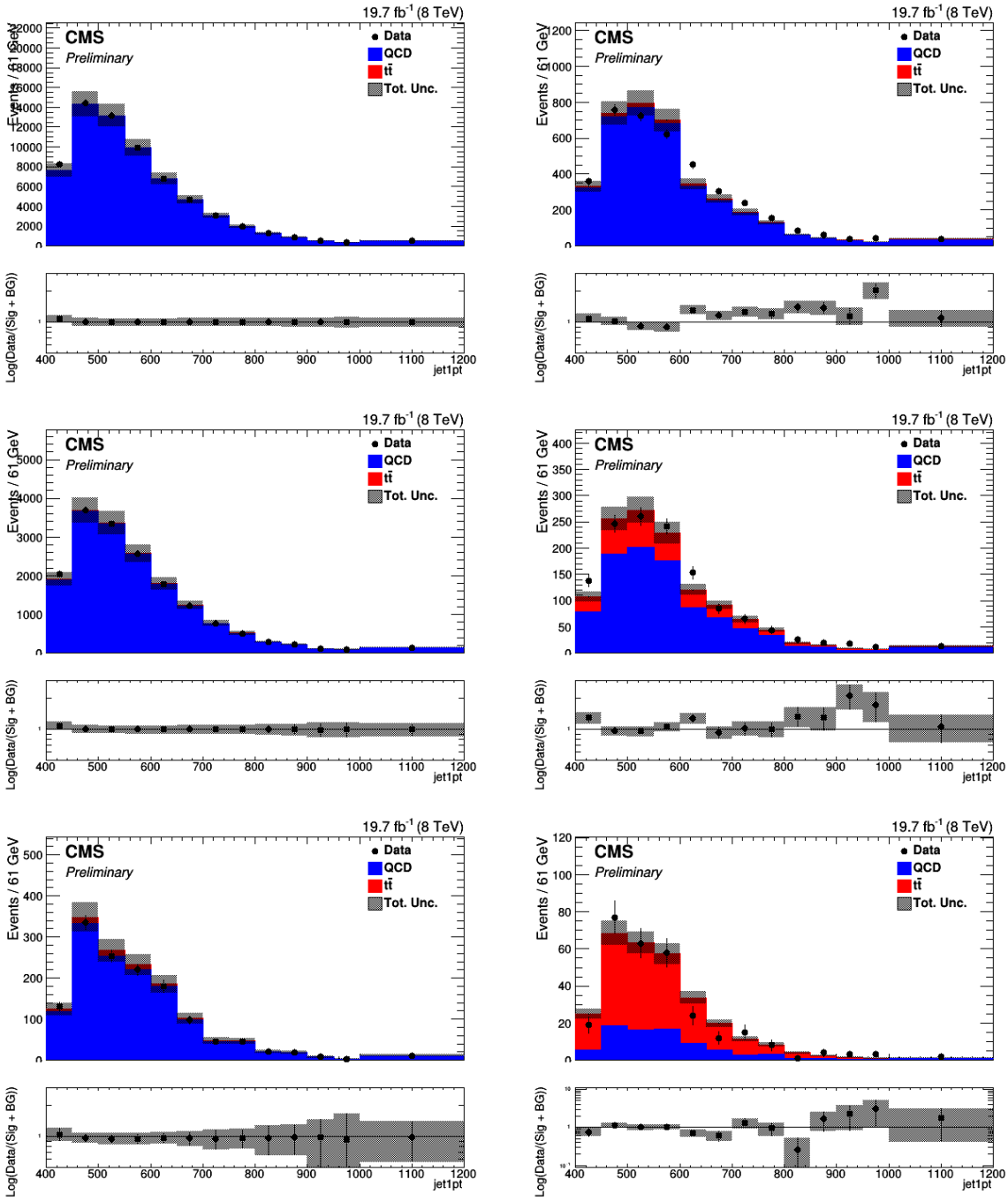

Figure 4.21: (Top Panels) Distribution of leading jet p_T in the antitag (left) and tag (right) region for events with 0 btags (top), 1 btag (middle) and 2 btags (bottom) for the inclusive ($p_T > 400$ GeV) sample. (Bottom Panels) Ratio of data to combined signal + background.

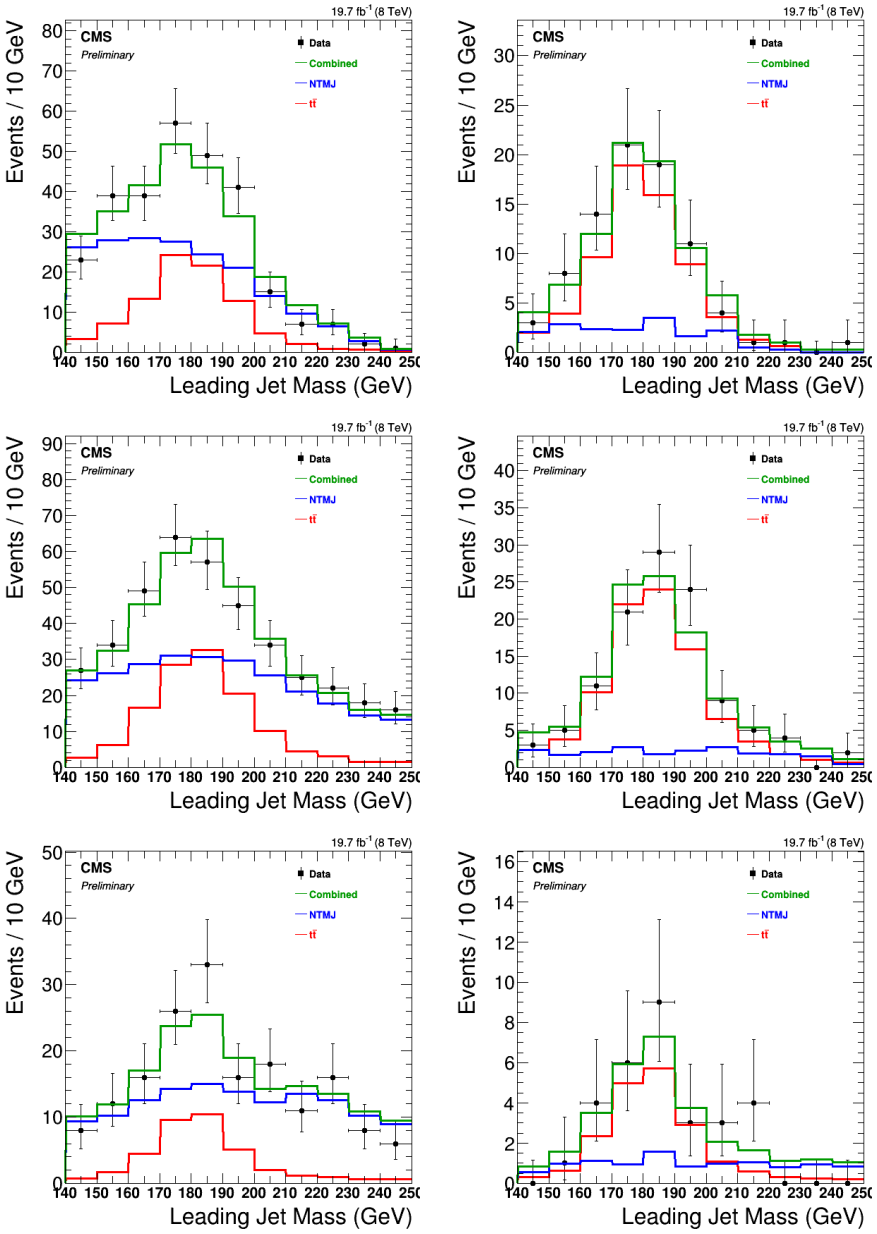

Figure 4.22: Postfit results of leading jet mass for (left) 1 btag and (right) 2 btags for (from top to bottom) 400< p_T <500, 500< p_T <600 and 600< p_T <700 GeV for the tagged region.

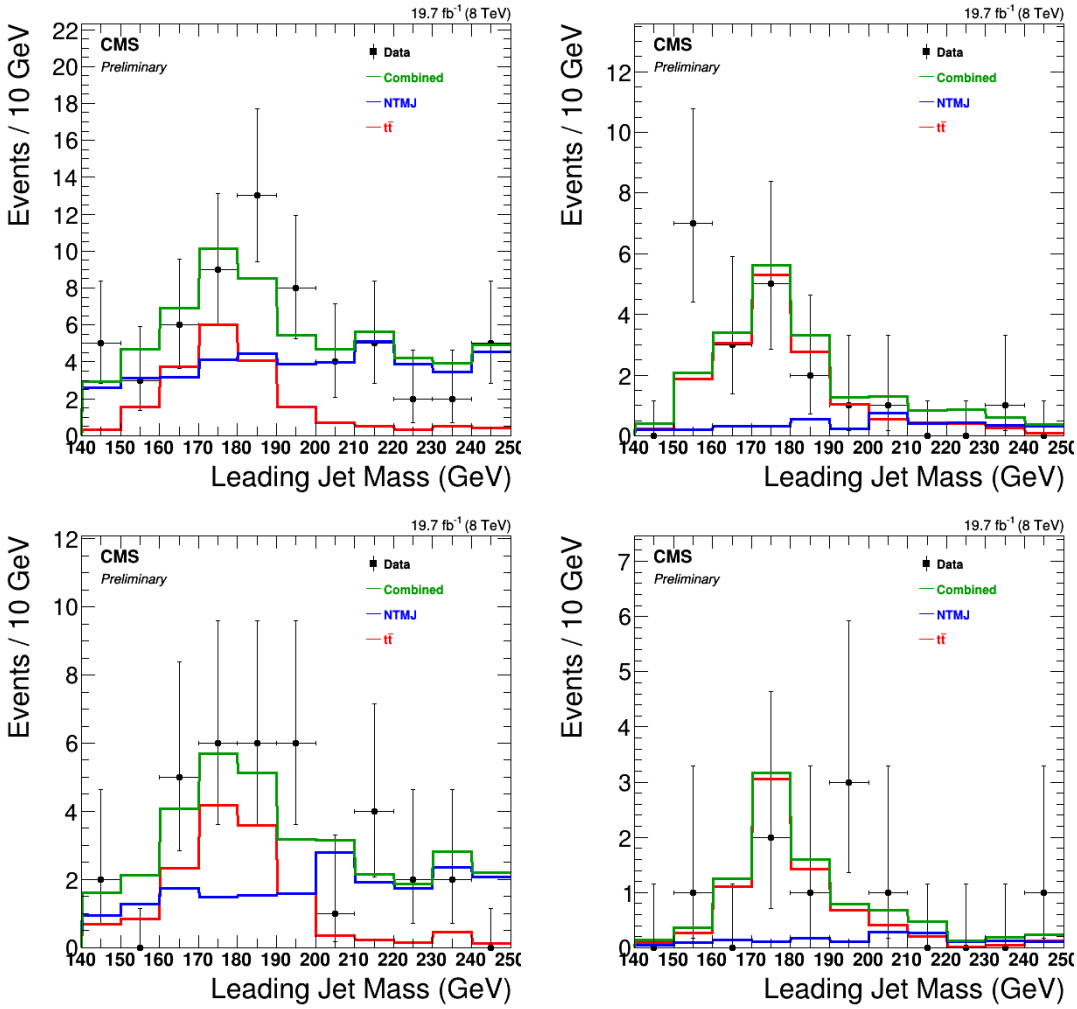

Figure 4.23: Postfit results of leading jet mass for (left) 1 btag and (right) 2 btags for (from top to bottom) 700< p_T <800 and 800< p_T <1200 GeV for the tagged region.

4.8 Unfolding

Unfolding relies on a response matrix to map reconstructed events to MC generated events. There are three conceivable outcomes when comparing the generated MC events to the reconstructed MC events: Pass—the generated top quark is part of a reconstructed top jet that passes the full reconstruction and analysis-level cuts, Fail—the generated top quark is part of a reconstructed top jet that fails any reconstruction or analysis-level cut, and Fake—a top jet that passes the full reconstruction and analysis-level cuts does not correspond to a generated $t\bar{t}$ that would have passed.

The response matrix created by following the unfolding procedure is shown in Fig. 4.24.

We take the results from the individual p_T fits and combine them into a single distribution and unfold it to get the corresponding *generated* (or *theoretical*) distribution. The p_T binning was chosen to match the boosted semileptonic analysis [32]. The SVD[3] method of unfolding requires a choice of regularization parameter (k_{reg}). This parameter was varied from 2 to nbins=5, and the optimal value was found to be $k_{reg} = 2$ by following the procedure in Ref. [33]. This procedure involves plotting the values of the "d" vector vs. k_{reg} (Fig. 4.26). The point at which this distribution becomes constant is the proper value for k_{reg}; in this case $k_{reg} = 2$.

The unfolding procedure was tested for bias by dividing the dataset into two parts and unfolding the first part with a response matrix generated by the second part. The

[3]Singular Value Decomposition in RooUnfold.

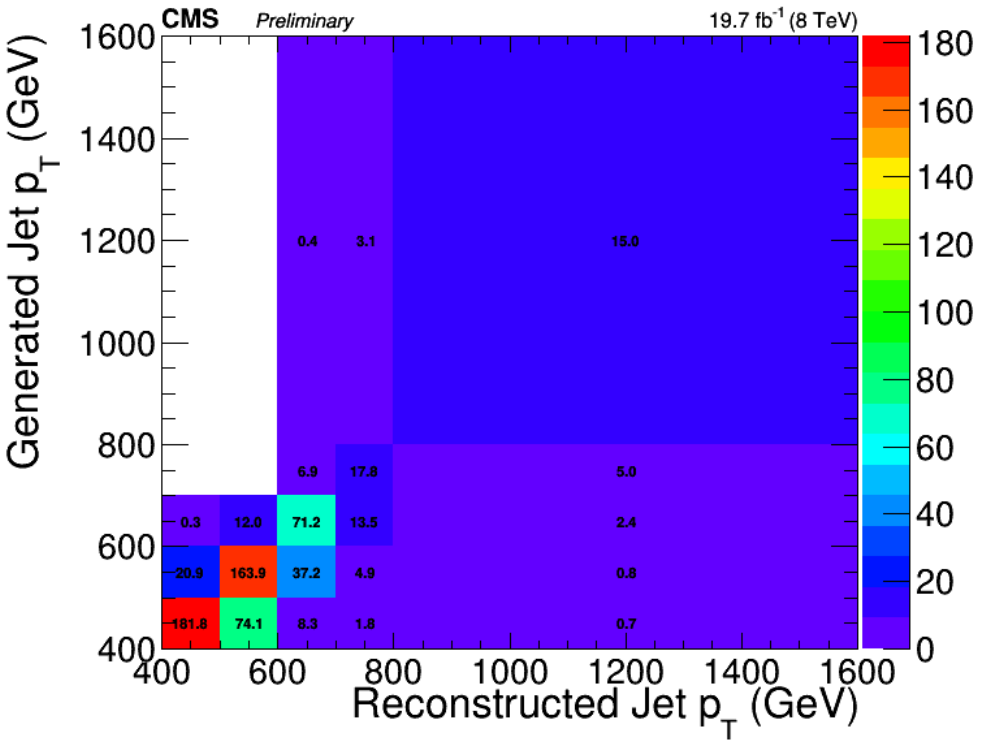

Figure 4.24: Plot of reponse matrix relating generated and reconstructed p_T.

test showed no bias and the result of it can be seen in Fig. 4.27.

A separate response matrix is created for a $\pm 1\sigma$ shift in each of the systematics in Sec. 4.6 that affect the shape of the top p_T distribution. The nominal fit result is then unfolded with each of these response matrices to characterize the effect of the systematic on the unfolding procedure. The difference between the unfolded result using the nominal response matrix and the unfolded result using a response matrix for shifted systematic is taken as the uncertainty in the unfolding due to that systematic effect. These are then added in quadrature with each other and the statistical error in the nominal result to give the result in Fig. 4.25.

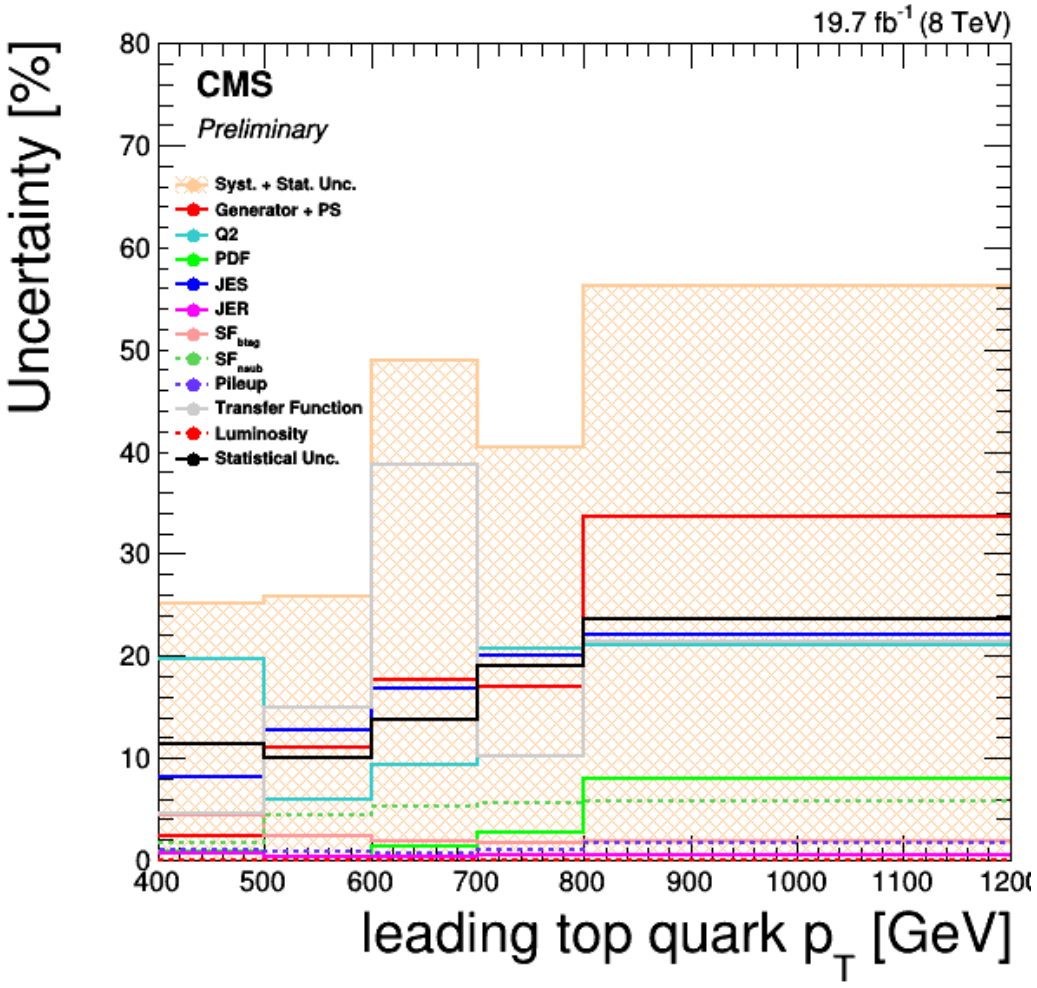

Figure 4.25: Plot of uncertainty due to systematic effects on the unfolded result.

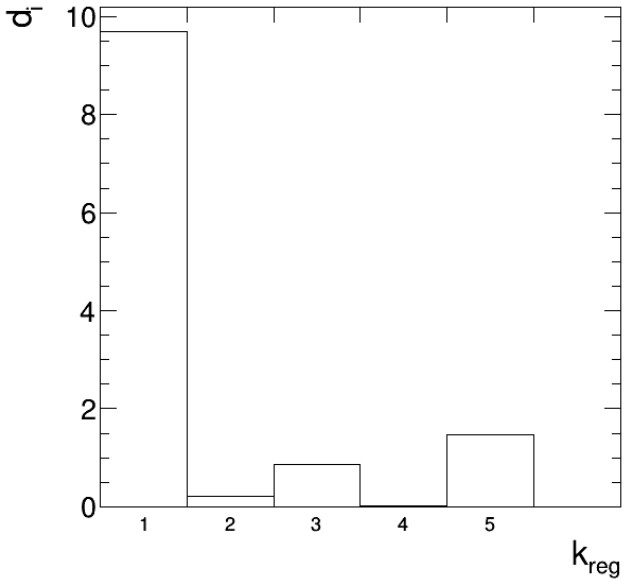

Figure 4.26: Plot of d_i vs k_{reg}. The point at which d_i starts to be constant ($k_{reg} = 2$) is the optimal k_{reg} for unfolding.

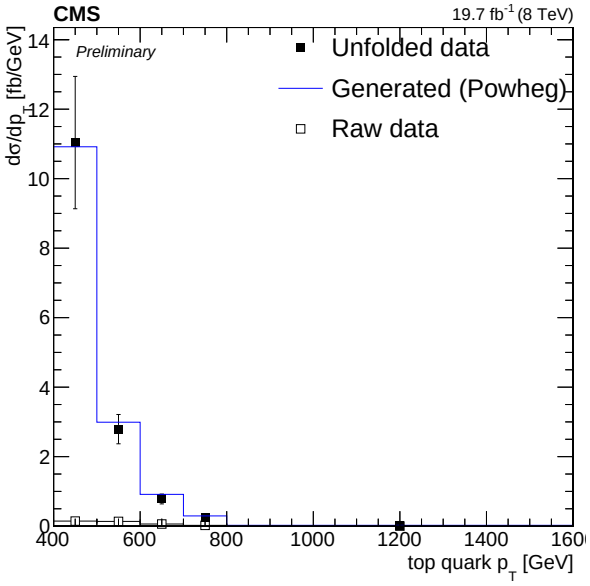

Figure 4.27: Bias test for unfolding procedure. Half of the data was unfolded with a response matrix generated from the other half.

4.9 Results

4.9.1 Inclusive Cross Section

The templates shown in the left-hand-side of Fig. 4.20 are fit to get the inclusive cross section. The result is shown in the right-hand-side of Fig. 4.20 and leads to the following cross section:

$$\sigma_{t\bar{t}} = 1.39 \pm 0.08 \text{ (stat)} \pm 0.09 \text{ (exp)} \pm 0.42 \text{ (theory) pb}$$

This result is approximately 16% smaller than the expected result for POWHEG, 1.66 pb.

4.9.2 Differential Cross Section

The differential cross section was discussed earlier in Sec. 1.3 and is given by Eq. 1.3. The unfolding procedure implicitly includes both the acceptance and selection efficiency, so they do not need to be calculated individually. Event counts:

p_T (GeV)	Data	Stat (%)	Exp (%)	Th (%)	Tot (%)	POWHEG	MC@NLO	Madgraph	SemiLeptonic
400-500	8.59	11.4	10.5	60.4	62.4	11.8	10.1	13.1	10.4
500-600	2.51	10.0	20.3	19.4	29.8	3.20	2.63	3.64	2.74
600-700	0.789	13.7	42.5	7.0	45.2	0.972	0.754	1.11	0.786
700-800	0.266	19.0	23.2	9.3	31.4	0.322	0.238	0.363	0.254
800-1200	0.042	23.5	31.3	19.7	43.8	0.049	0.030	0.050	0.036

Table 4.6: Differential $t\bar{t}$ cross section as a function of $p_T(\frac{d\sigma}{dp_T})(fb\text{GeV}^{-1})$ unfolded to parton-level with uncertainty. The Data column is the result of this analysis. Also compared are three generator and parton-shower combinations: POWHEG+Pythia6, MC@NLO+Herwig6, and Madgraph+Pythia6. The uncertainty is divided into Statistical, Experimental, Theoretical, and Total as described in Section 4.6.

The parton-level unfolded differential cross section is compared to different generators as well as the CMS 8 TeV semileptonic result and shown in Fig. 4.28.

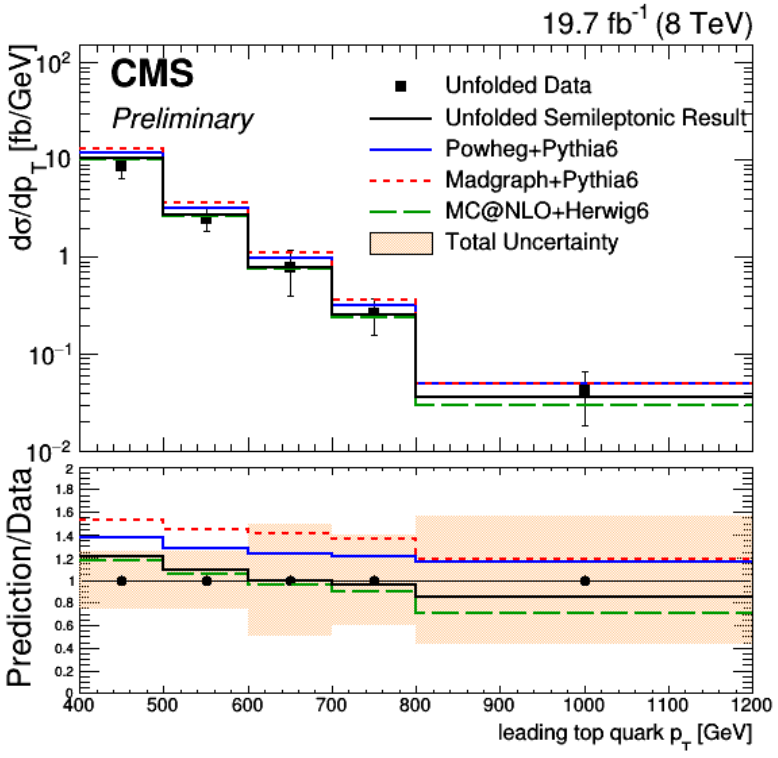

Figure 4.28: Parton-level differential $t\bar{t}$ cross section.

There is good agreement between the unfolded result and the theoretical input to Powheg.

Chapter 5

Summary

The differential $t\bar{t}$ cross section in the highly boosted all-hadronic channel at $\sqrt{s} =$ 8TeV at the LHC has been presented. This measurement required understanding a data-driven background that was correlated to one of the primary discriminants of signal from background, τ_{32}. The Alphabet method was used to overcome this difficulty by using the sidebands to interpolate a curve to account for the correlation. The combination of event selection and jet collection used were unique for the all-hadronic channel, and as a result, data/MC scale factors for n-subjettiness and subjet btagging were measured and used.

A simultaneous fit of $t\bar{t}$ Monte Carlo and a data-driven background for QCD was performed in bins of subjet btags. This fit was repeated for the p_T ranges 400-500GeV, 500-600GeV, 600-700GeV, 700-800GeV, and 800-1600GeV. The fits accounted for systematics including but not limited to jet energy corrections, scale factor errors,

and errors in the alphabet-based background measurement. The results of the fit were unfolded by singular value decomposition in order to compare them with theory. The result is shown in Fig. 4.28. The result found to be in with NLO+PS simulation calculations. The inclusive integrated $t\bar{t}$ cross section for jets with $p_T > 400$ GeV was also measured, and the result of $\sigma_{t\bar{t}} = 1.39 \pm 0.08$ (stat)$\pm$0.09 (exp)$\pm$0.42 (theory) pb slightly smaller than the POWHEG expectation of $\sigma_{t\bar{t}} = 1.66$ pb, but agrees within uncertainties.

Chapter 6

Appendix

6.1 Additional Figures and Details

6.1.1 Background Fit

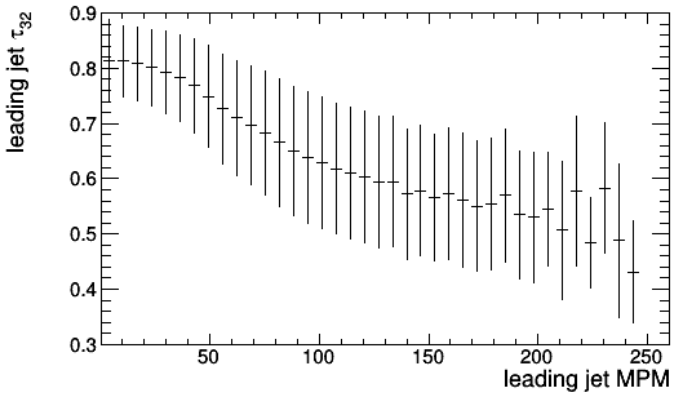

Figure 6.1: Profile histogram between τ_{32} and Minimum Pairwise Mass (MPM) demonstrating the correlation. Error bars are RMS.

6.1.2 Fit

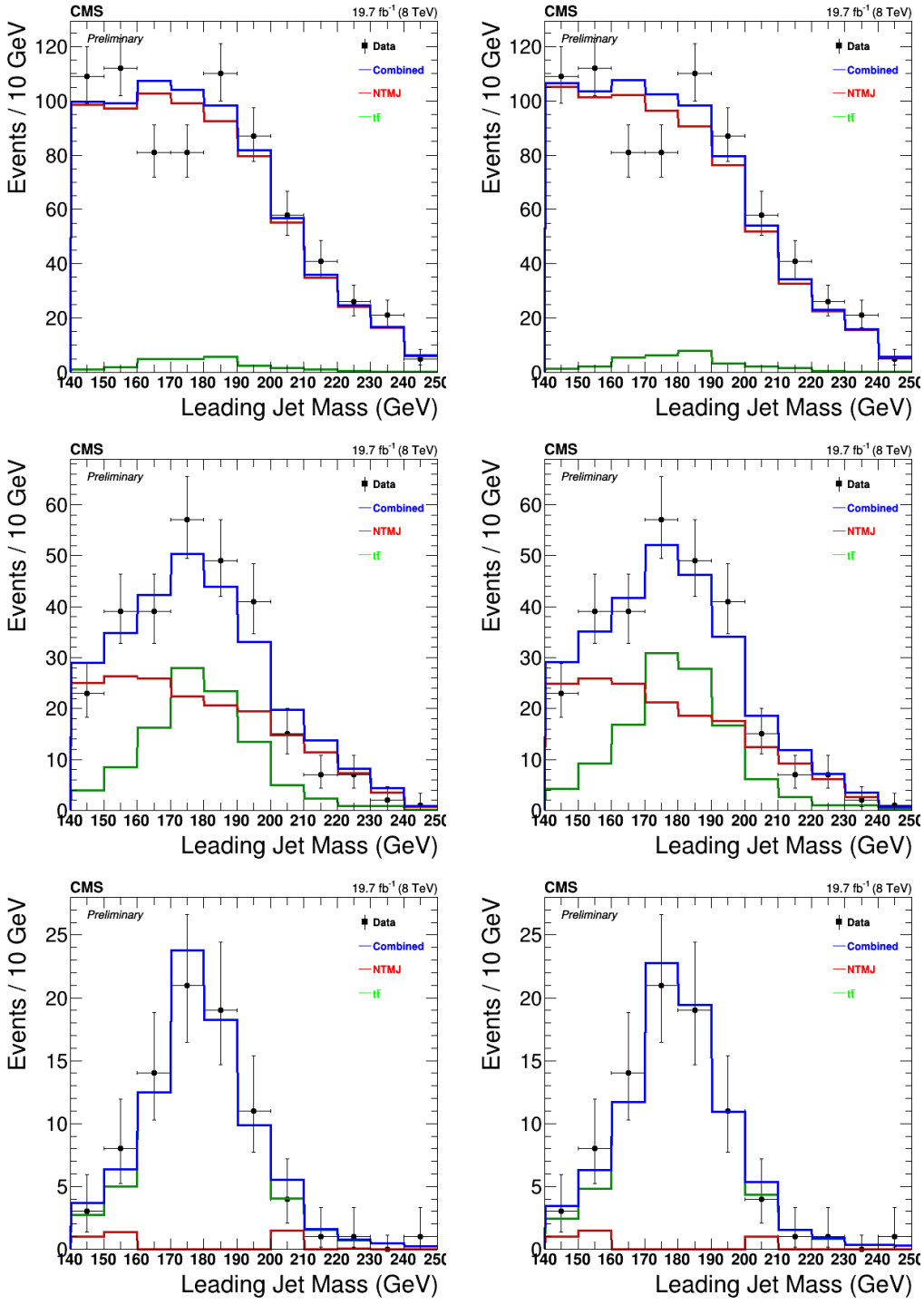

Figure 6.2: Prefit(left) and Fit(right) of 0 btag, 1 btag, 2 btag for 400< p_T <500 GeV.

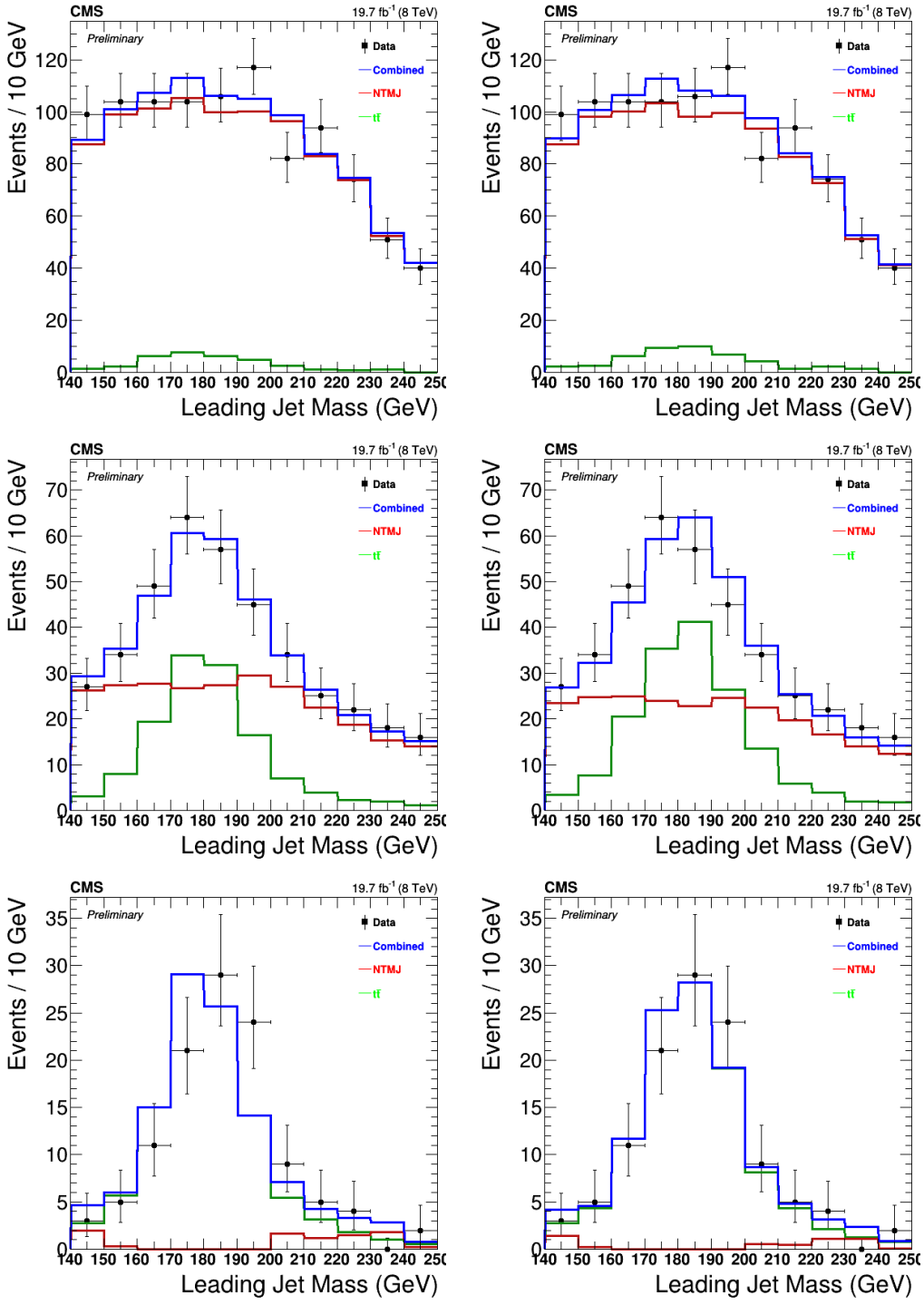

Figure 6.3: Prefit(left) and Fit(right) of 0 btag, 1 btag, 2 btag for 500< p_T <600 GeV.

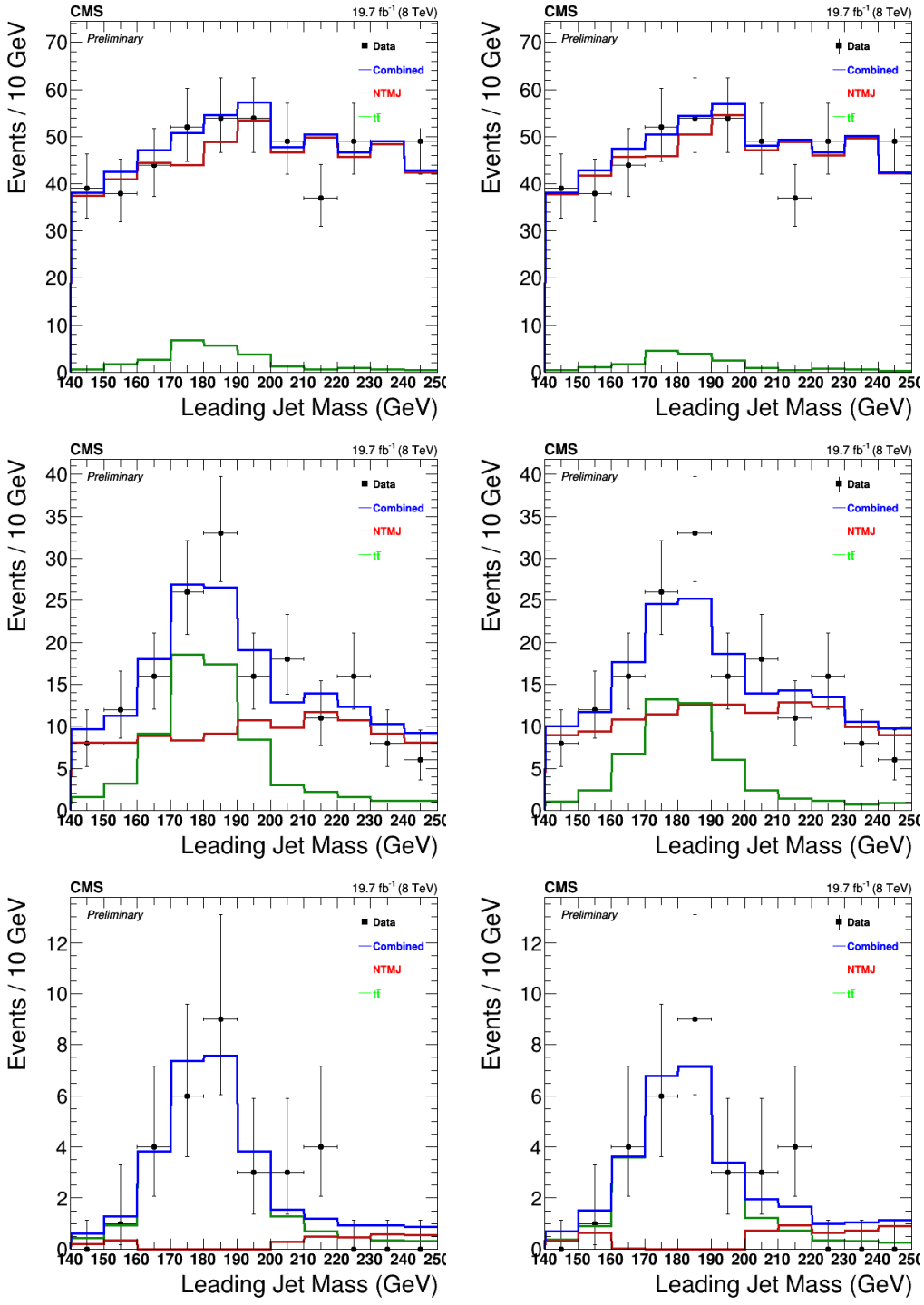

Figure 6.4: Prefit(left) and Fit(right) of 0 btag, 1 btag, 2 btag for 600< p_T <700 GeV.

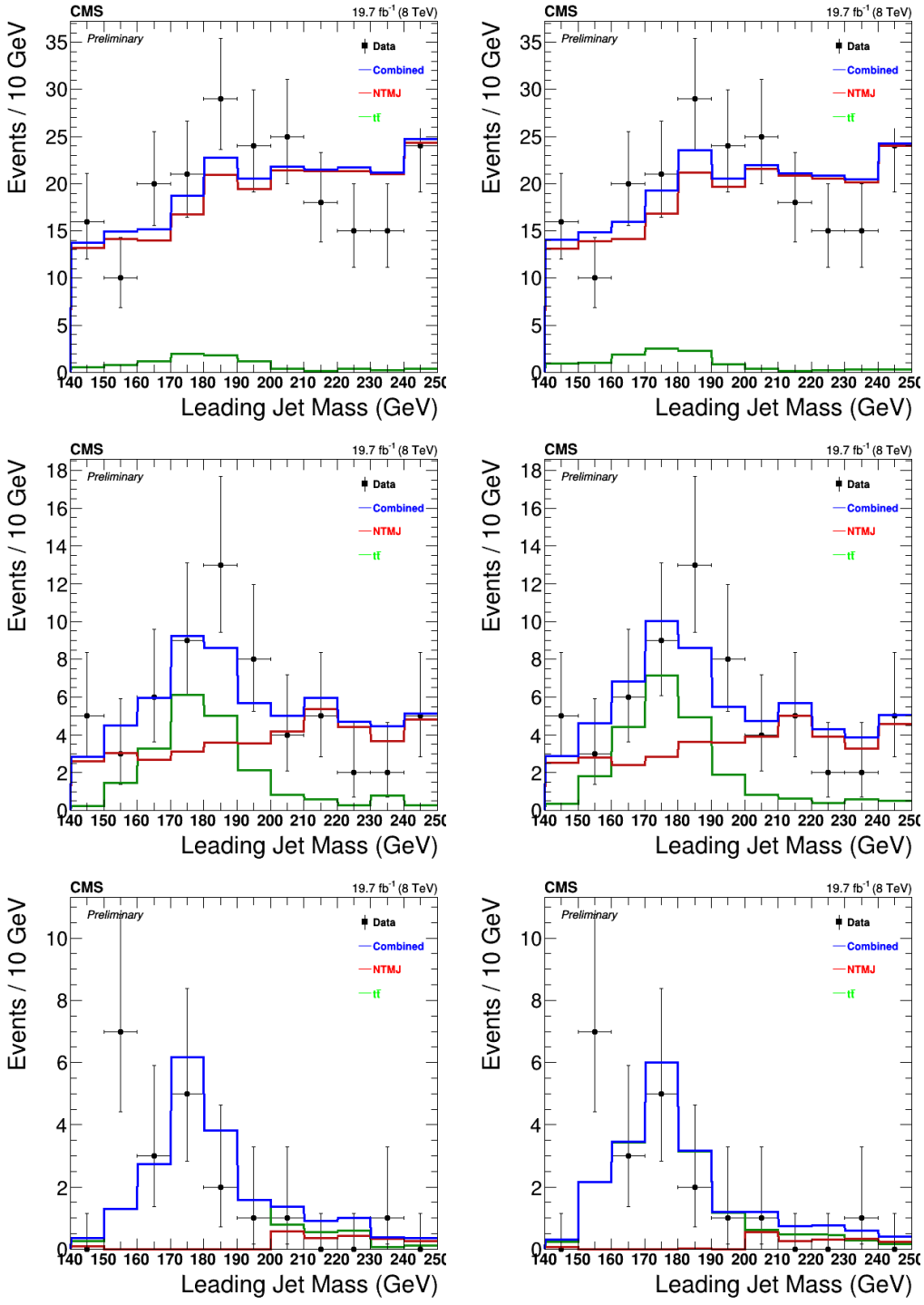

Figure 6.5: Prefit(left) and Fit(right) of 0 btag, 1 btag, 2 btag for 700< p_T <800 GeV.

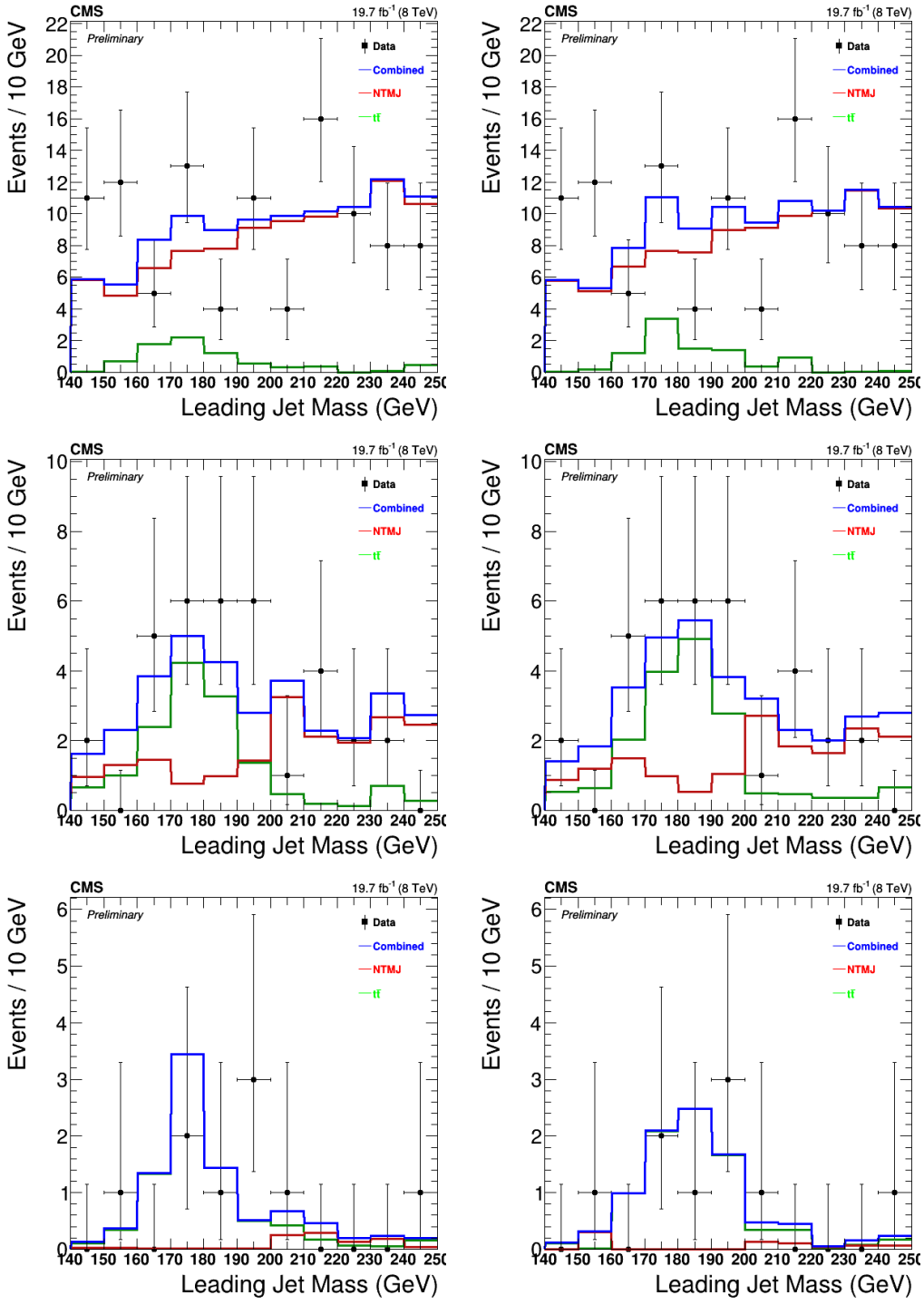

Figure 6.6: Prefit(left) and Fit(right) of 0 btag, 1 btag, 2 btag for 800< p_T <1600 GeV.

6.1.3 Post-fit Kinematics

Jet parameters for both the leading and subleading jet will be shown post-fit in the following figures.

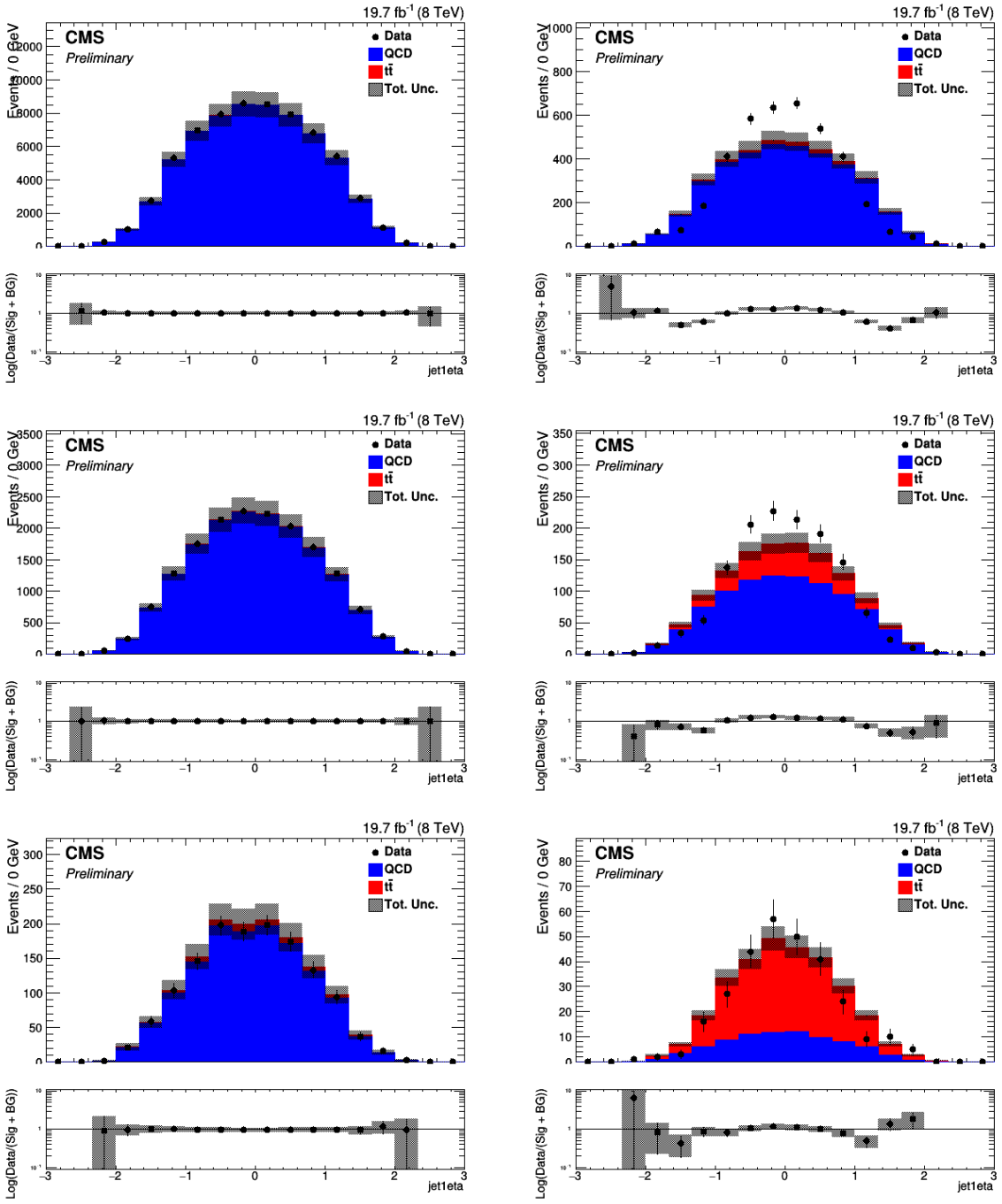

Figure 6.7: Pseudorapidity distribution of the leading jets in the antitag (left) and the tag (right) region for 0 btags (top), 1 btag (middle) and 2 btags (bottom) after selection, but before after selection and fitting. The discrepancy of the leading jet η is well understood and comes from the τ_{32} requirement.

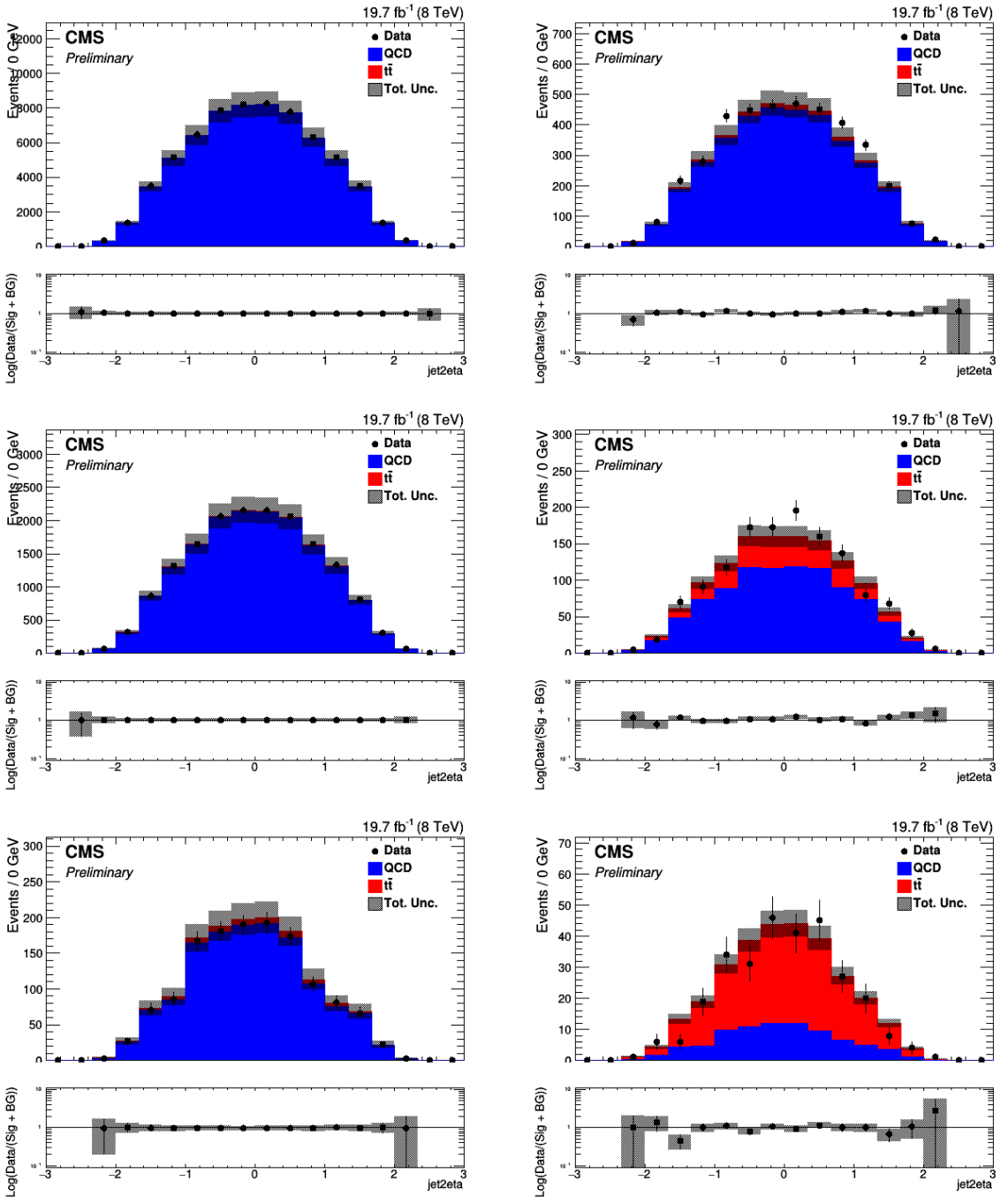

Figure 6.8: Pseudorapidity distribution of the subleading jets in the antitag (left) and the tag (right) region for 0 btags (top), 1 btag (middle) and 2 btags (bottom) after selection and fitting.

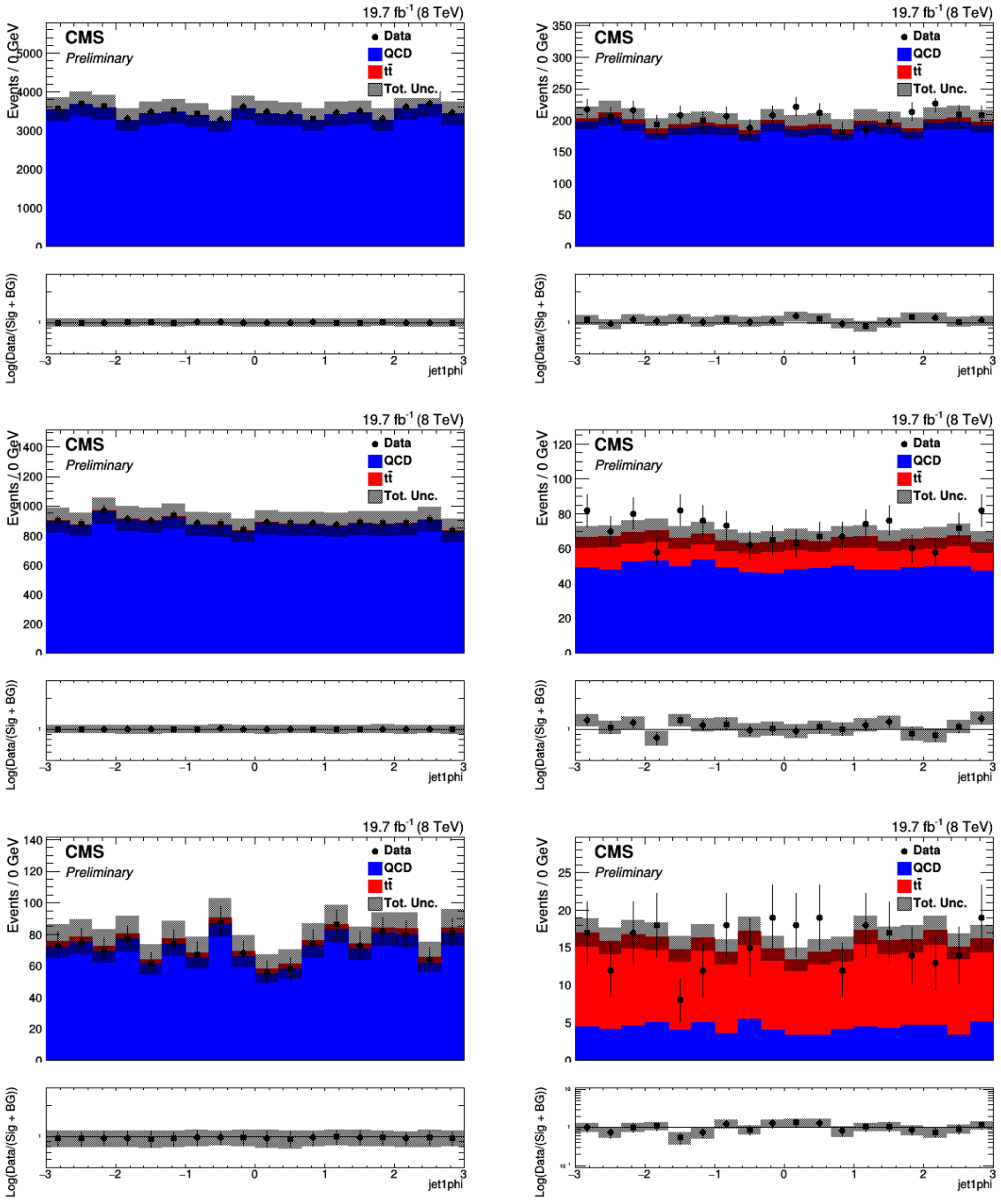

Figure 6.9: Phi distribution of the leading jets in the antitag (left) and the tag (right) region for 0 btags (top), 1 btag (middle) and 2 btags (bottom) after selection and fitting.

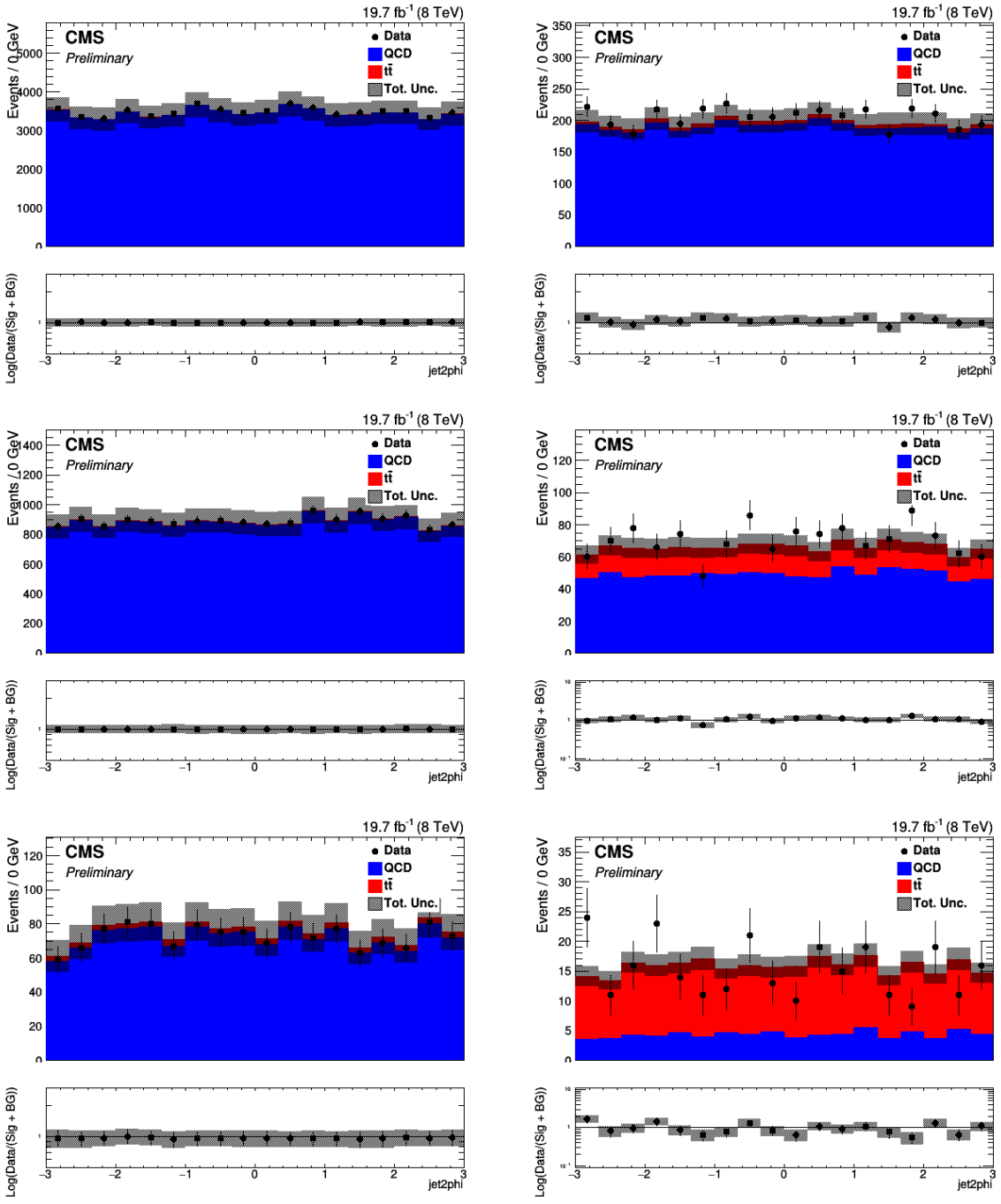

Figure 6.10: Phi distribution of the subleading jets in the antitag (left) and the tag (right) region for 0 btags (top), 1 btag (middle) and 2 btags (bottom) after selection and fitting.

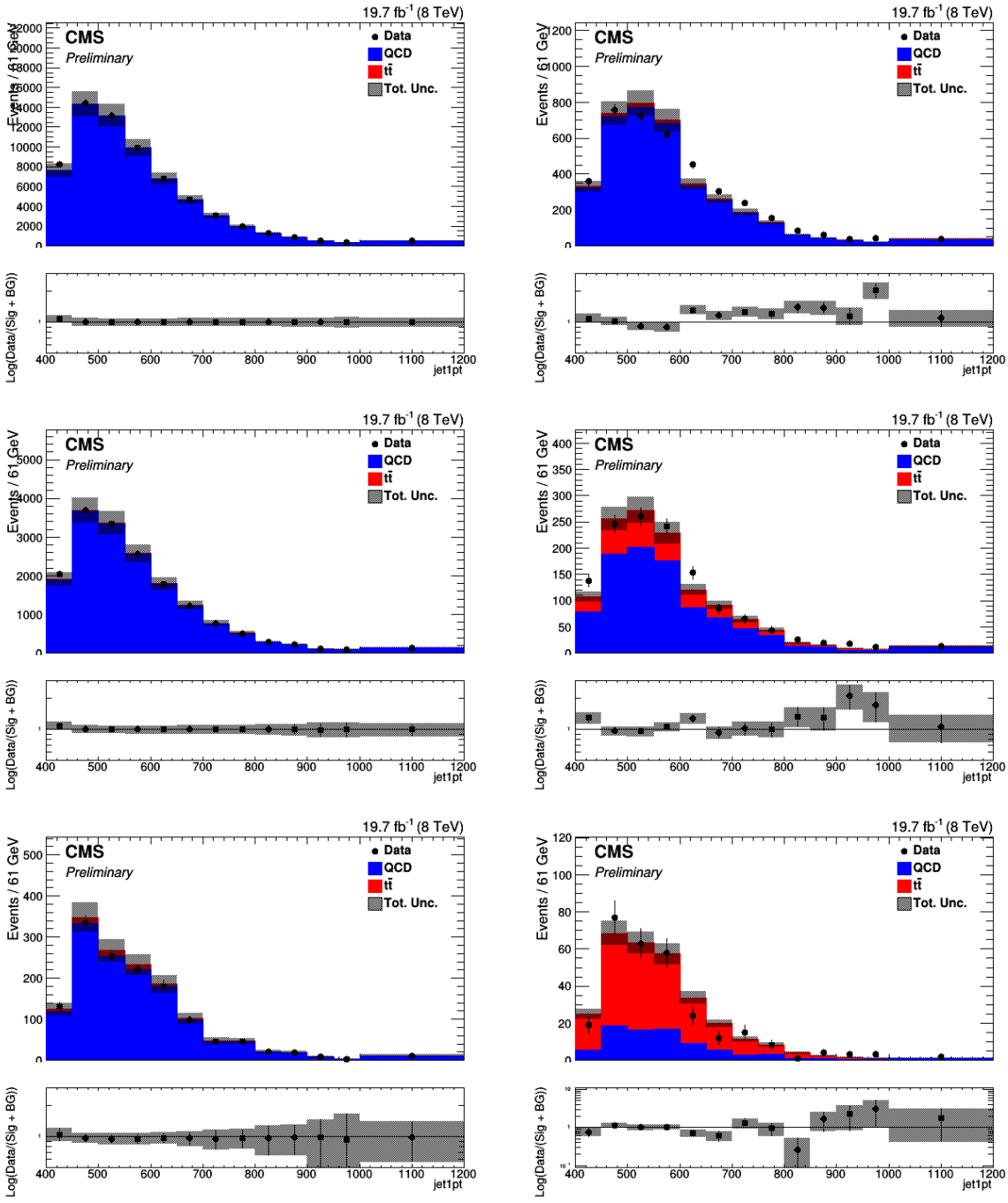

Figure 6.11: Transverse momentum distribution of the leading jets in the antitag (left) and the tag (right) region for 0 btags (top), 1 btag (middle) and 2 btags (bottom) after selection and fitting.

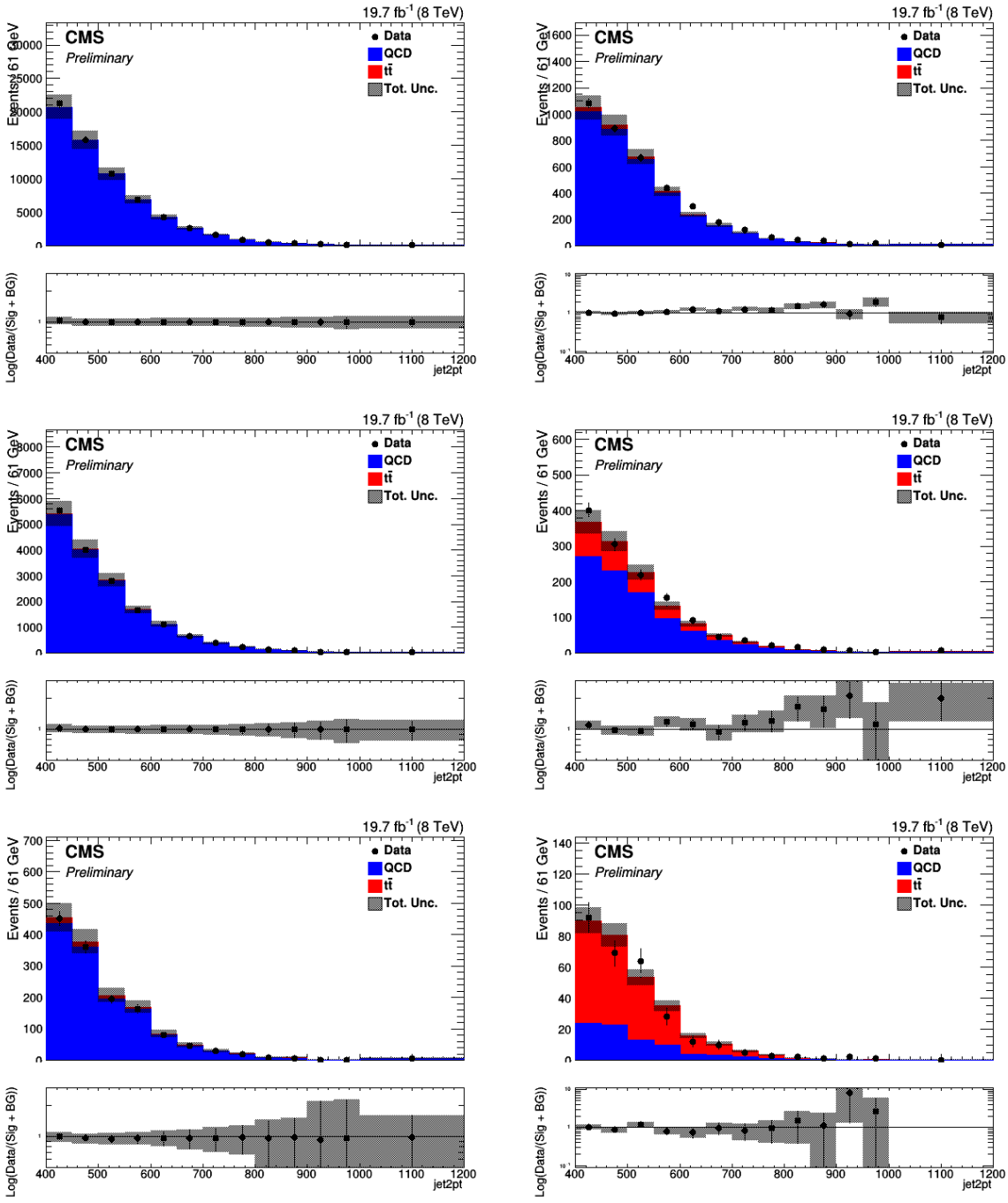

Figure 6.12: Transverse momentum distribution of the subleading jets in the antitag (left) and the tag (right) region for 0 btags (top), 1 btag (middle) and 2 btags (bottom) after selection and fitting.

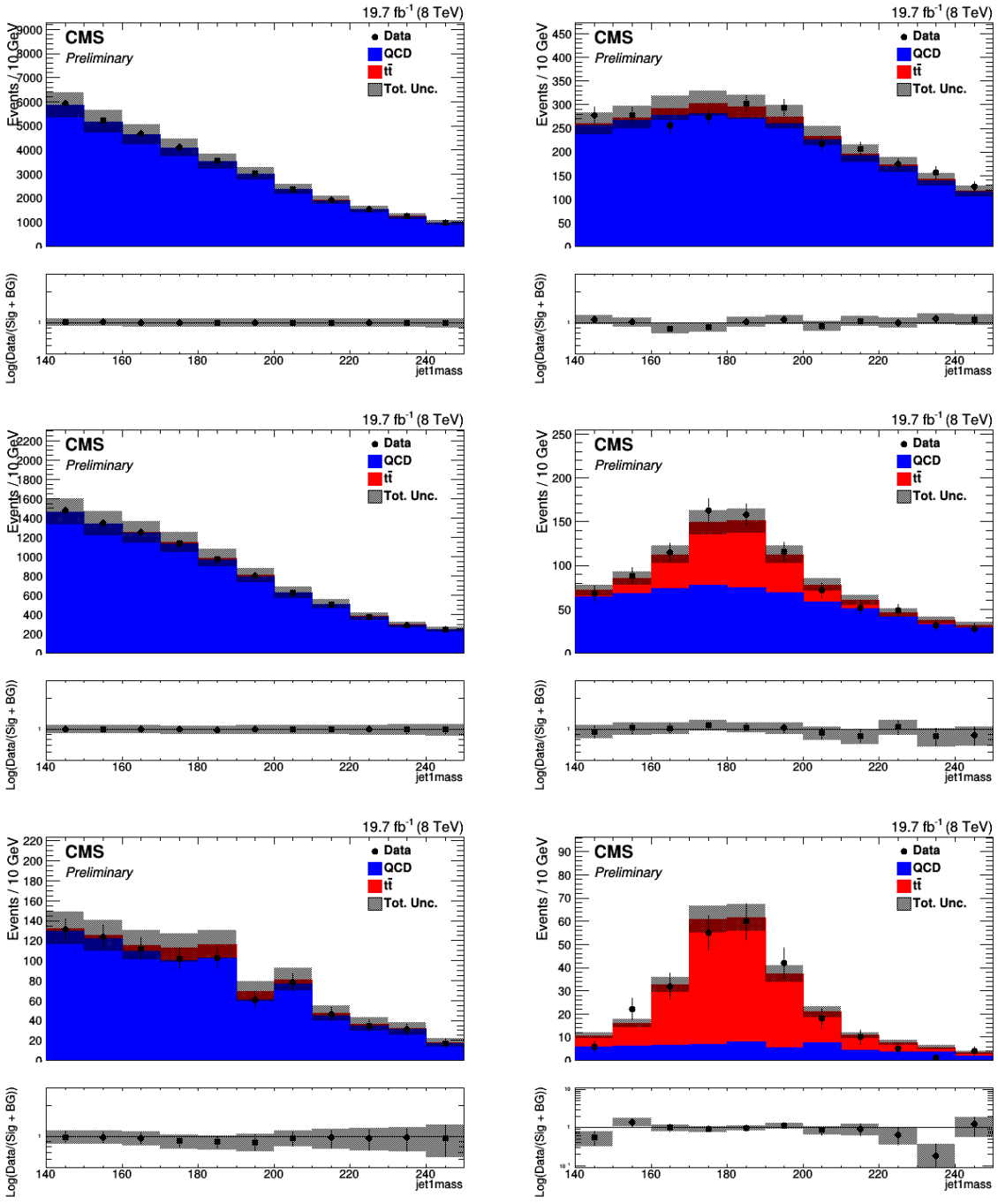

Figure 6.13: Mass distribution of the leading jets in the antitag (left) and the tag (right) region for 0 btags (top), 1 btag (middle) and 2 btags (bottom) after selection and fitting.

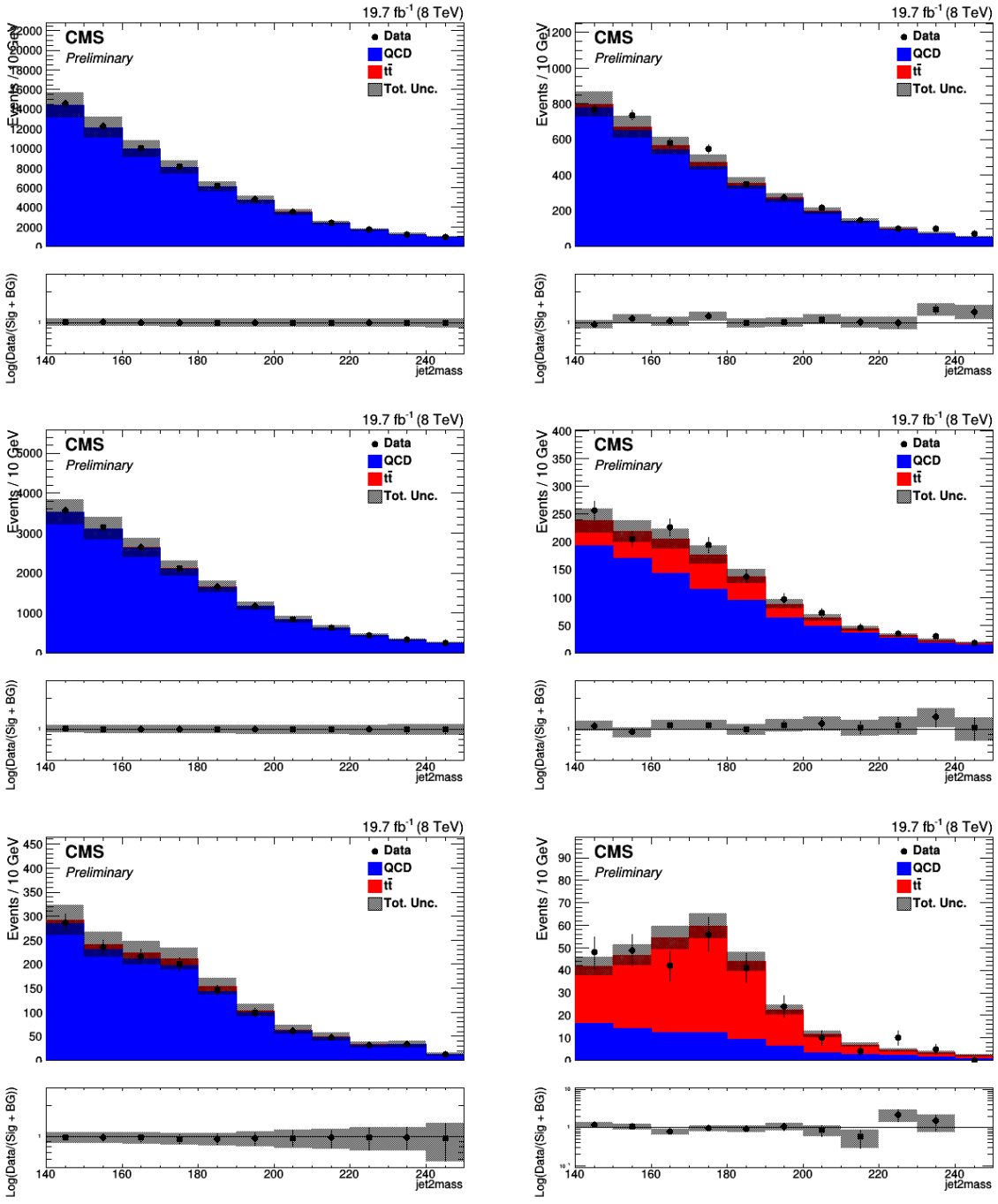

Figure 6.14: Mass distribution of the subleading jets in the antitag (left) and the tag (right) region for 0 btags (top), 1 btag (middle) and 2 btags (bottom) after selection and fitting.

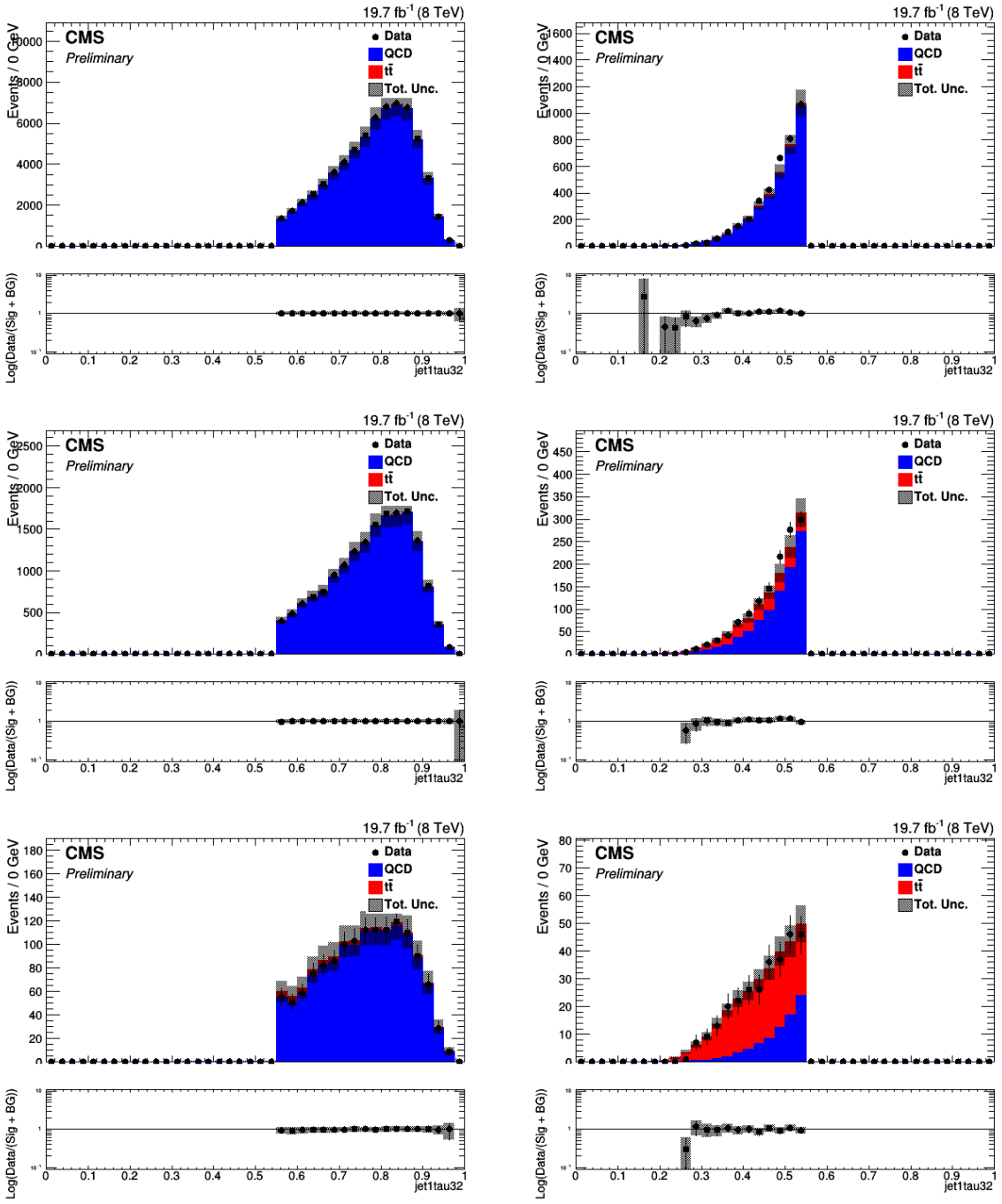

Figure 6.15: τ_{32} distribution of the leading jets in the antitag (left) and the tag (right) region for 0 btags (top), 1 btag (middle) and 2 btags (bottom) after selection and fitting. For the tagged distribution, the normalization for NTMJ is taken from the sidebands.

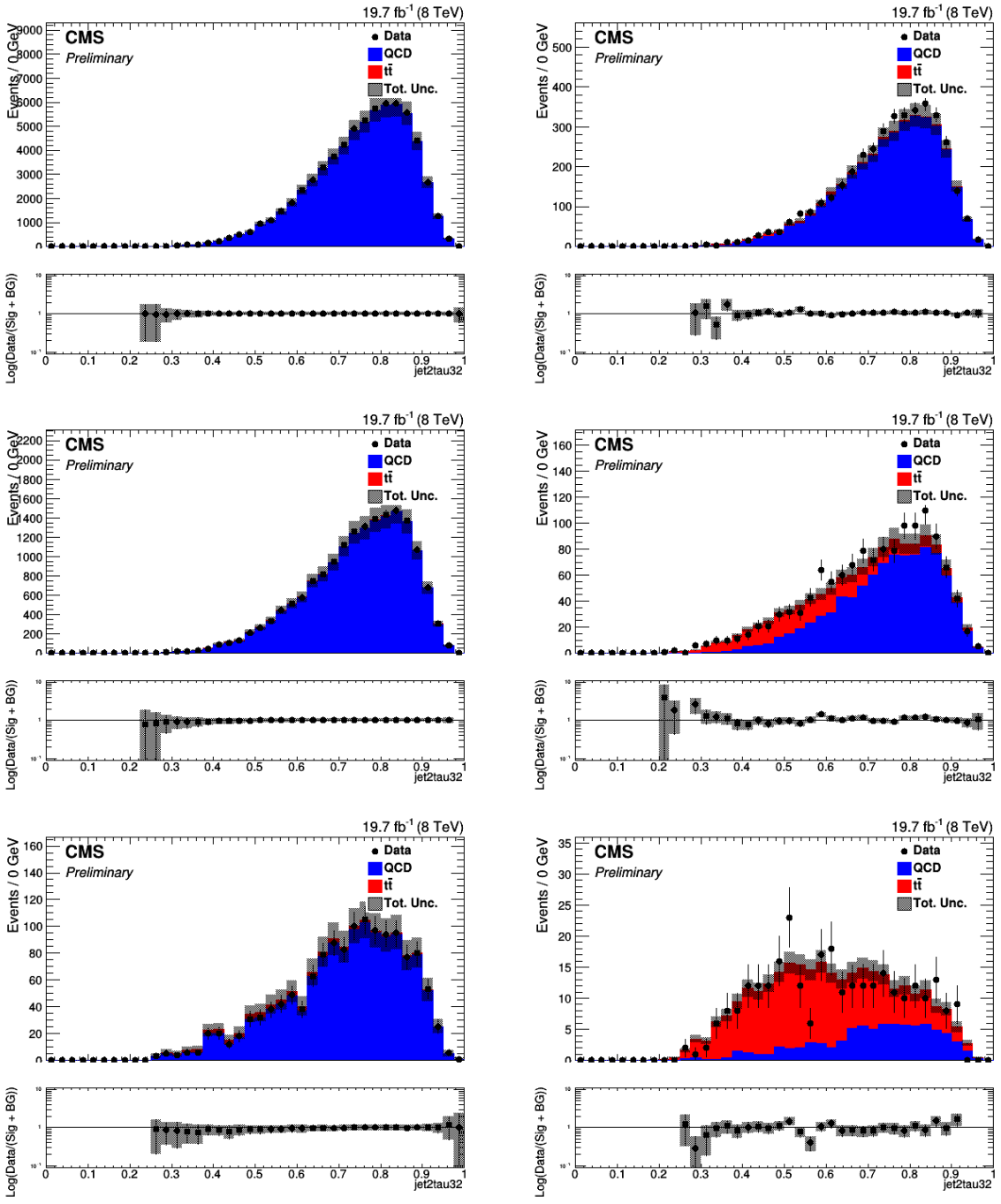

Figure 6.16: τ_{32} distribution of the subleading jets in the antitag (left) and the tag (right) region for 0 btags (top), 1 btag (middle) and 2 btags (bottom) after selection and fitting.

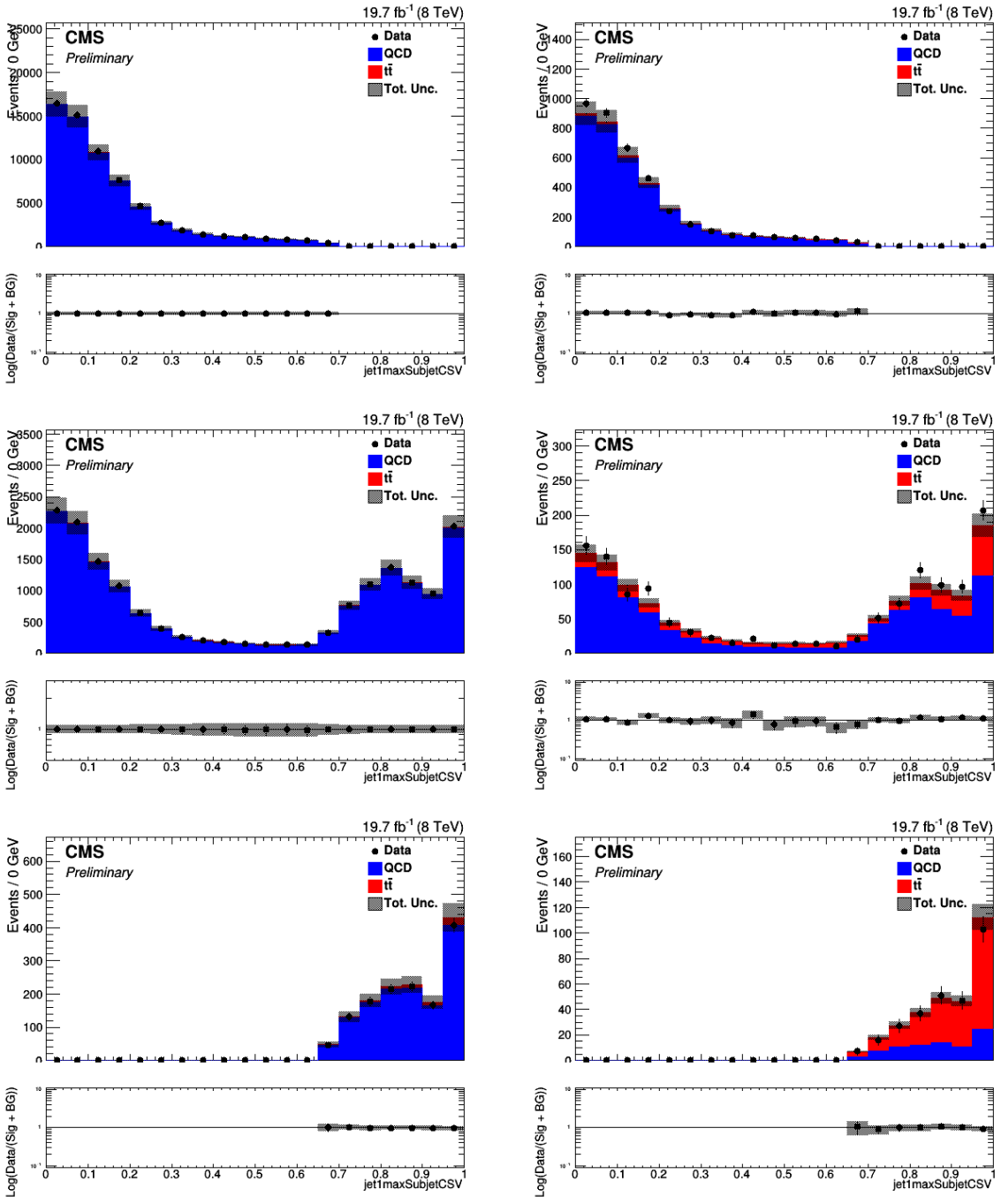

Figure 6.17: Maximum CSV from all subjets of the leading jets in the antitag (left) and the tag (right) region for 0 btags (top), 1 btag (middle) and 2 btags (bottom) after selection and fitting.

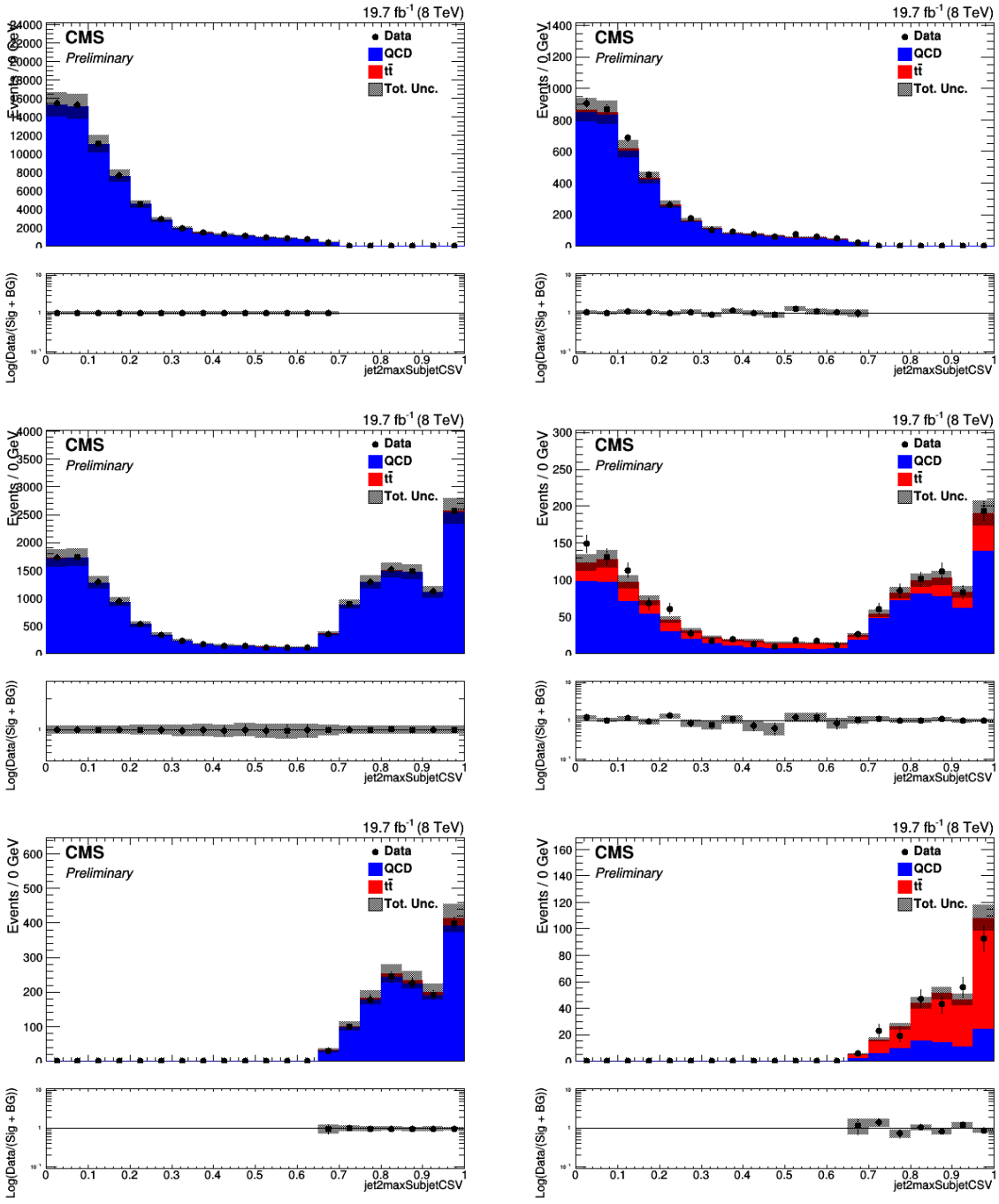

Figure 6.18: Maximum CSV from all subjets of the subleading jets in the antitag (left) and the tag (right) region for 0 btags (top), 1 btag (middle) and 2 btags (bottom) after selection and fitting.

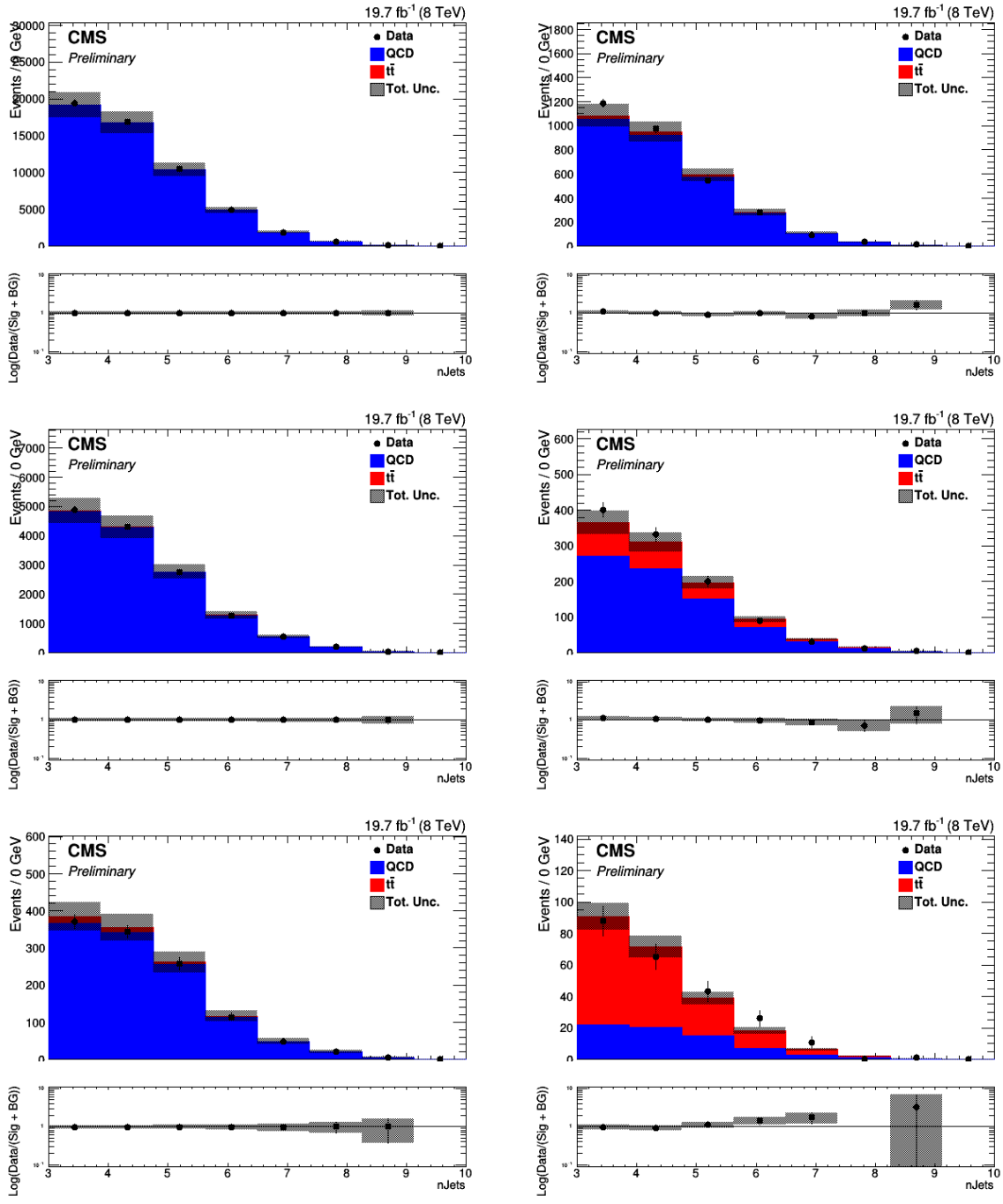

Figure 6.19: Number of jets spectrum in the antitag (left) and the tag (right) region for 0 btags (top), 1 btag (middle) and 2 btags (bottom) after selection and fitting.

Bibliography

[1] "Standard model of particle physics infographic," https://cds.cern.ch/journal/CERNBulletin/2012/35/News

[2] "Martin-stirling-thorne-watt parton distribution functions," https://mstwpdf.hepforge.org/.

[3] S. Moch, "Expectations at LHC from hard QCD," *J. Phys.*, vol. G35, p. 073001, 2008.

[4] N. Kidonakis, "Theoretical results for top-quark cross sections and distributions," in *38th International Conference on High Energy Physics (ICHEP 2016) Chicago, IL, USA, August 03-10, 2016*, 2016. [Online]. Available: http://inspirehep.net/record/1487920/files/arXiv:1609.07404.pdf

[5] J.-L. Caron, "Layout of the LEP tunnel including future LHC infrastructures." Feb 1997, aC Collection. Legacy of AC. Pictures from 1992 to 2002. [Online]. Available: http://cds.cern.ch/record/841560

BIBLIOGRAPHY

[6] "Cross-sectional view of cms detector," http://cds.cern.ch/record/2205172.

[7] "View of the cms silicon tracker," http://accms04.physik.rwth-aachen.de/ schael/Stefan

[8] "Schematic of silicon pixel," http://cms.web.cern.ch/news/silicon-pixels.

[9] "The cms electromagenetic calorimeter," http://cms.web.cern.ch/news/electromagnetic-calorimeter.

[10] "The cms hadron calorimeter," http://cms.web.cern.ch/news/cms-prepares-pixel-and-hcal-upgrades.

[11] "Muon drift tubes and iron return yoke," https://dorigo.wordpress.com/2007/04/.

[12] "Missing et schematic diagram," https://cms-docdb.cern.ch/cgi-bin/PublicDocDB/ShowDocument?docid=12312.

[13] M. Cacciari, G. P. Salam, and G. Soyez, "The Anti-k(t) jet clustering algorithm," *JHEP*, vol. 04, p. 063, 2008.

[14] D. J. Griffiths, *Introduction to elementary particles; 2nd rev. version*, ser. Physics textbook. New York, NY: Wiley, 2008. [Online]. Available: https://cds.cern.ch/record/111880

[15] "The ckm quark-mixing matrix," http://pdg.lbl.gov/2015/reviews/rpp2015-rev-ckm-matrix.pdf.

[16] "Cms physics: Technical design report volume 1: Detector performance and software," http://cds.cern.ch/record/922757.

[17] S. Chatrchyan *et al.*, "Description and performance of track and primary-vertex reconstruction with the CMS tracker," *JINST*, vol. 9, no. 10, p. P10009, 2014.

[18] "Passage of particles through matter," http://pdg.lbl.gov/2011/reviews/rpp2011-rev-passage-particles-matter.pdf.

[19] CMS Collaboration, "A cambridge-aachen (c-a) based jet algorithm for boosted top-jet tagging," CMS Physics Analysis Summary CMS-PAS-JME-009-01, 2009. [Online]. Available: https://cds.cern.ch/record/1194489

[20] J. Thaler and K. Van Tilburg, "Identifying Boosted Objects with N-subjettiness," *JHEP*, vol. 03, p. 015, 2011.

[21] CMS Collaboration, "Cms luminosity based on pixel cluster counting - Summer 2013 update," CMS Physics Analysis Summary CMS-PAS-LUM-13-001, 2013. [Online]. Available: http://cds.cern.ch/record/1598864

[22] ——, "Particle-flow event reconstruction in cms and performance for jets, taus, and met," CMS Physics Analysis Summary CMS-PAS-JME-009-01, 2009. [Online]. Available: https://cds.cern.ch/record/1194487

[23] S. Alioli, P. Nason, C. Oleari, and E. Re, "A general framework for implement-

ing NLO calculations in shower Monte Carlo programs: the POWHEG BOX," *JHEP*, vol. 06, p. 043, 2010.

[24] T. Sjöstrand, S. Mrenna, and P. Skands, "PYTHIA 6.4 physics and manual," *JHEP*, vol. 05, p. 026, 2006.

[25] S. Agostinelli *et al.*, "Geant4—a simulation toolkit," *Nucl. Instrum. Meth. A*, vol. 506, p. 250, 2003.

[26] J. Conway, J. Dolen, R. Erbacher, P. Maksimovic, K. Nash, M. Osherson, J. Pilot, S. Rappoccio, and R. Vasquez-Sierra, "Search for BSM $t\bar{t}$ production in the boosted all-hadronic final state using pp collisions at $\sqrt{s} = 8$ tev," CMS Note 2012/179, 2013.

[27] M. Cacciari, G. P. Salam, and G. Soyez, "The Catchment Area of Jets," *JHEP*, vol. 04, p. 005, 2008.

[28] S. Chatrchyan *et al.*, "Determination of Jet Energy Calibration and Transverse Momentum Resolution in CMS," *JINST*, vol. 6, p. P11002, 2011.

[29] J. Conway, J. Dolen, R. Erbacher, B. Hegner, G. Hu, J. Ott, P. Maksimovic, S. Rappoccio, and R. Vasquez-Sierra, "Search for BSM $t\bar{t}$ production in the boosted all-hadronic final state," CMS Note 2011/194, 2011.

[30] M. Botje, J. Butterworth, A. Cooper-Sarkar, A. de Roeck, J. Feltesse *et al.*, "The PDF4LHC Working Group Interim Recommendations," 2011.

[31] R. Barlow and C. Beeston, "Fitting using finite Monte Carlo samples," *Computer Physics Communications*, vol. 77, p. 219, 1993.

[32] M. Alyari, M. Bellis, S. Dittmer, S. Rappoccio, L. Skinnari, and J. Thom, "Measurement of the differential $t\bar{t}$ production cross section for high-p_T top quarks in e/μ+jets final states at 8 tev," CMS Note 2013/343, 2012.

[33] A. Hocker and V. Kartvelishvili, "Svd approach to data unfolding," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 372, no. 3, pp. 469 – 481, 1996. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0168900295014780

Vita

Education

Ph.D. Physics, Johns Hopkins University, Expected 2017.

M.A. Physics, Johns Hopkins University, 2012.

B.S. Physics, University of West Florida, 2005.

B.S. Chemistry, University of West Florida, 2005.

Analysis Skills

Computer Programming

C++

Python

R

ROOT

VITA

RooFit

SQL

UNIX shell scripting

Statistical Background

Numerical Methods

Error Analysis

Model Building

Linear Regression

Monte Carlo Methods

Probability Distributions

Likelihood Analysis

Mathematical Background

Multivariate Calculus

Linear Algebra

Matrix Factorization

Fourier Analysis

VITA

Selected Publications

CMS Collaboration. Measurement of the differential top-quark pair-production production cross section with boosted tops in the all-hadronic channel at $\sqrt{s}$=8 TeV. In progress.

CMS Collaboration. Measurement of the production cross section in pp collisions at with lepton + jets final states. Physics Letters B. 720:83-104. March 2013

Swartz, M.; **Fehling, D.**; et al. A new technique for the reconstruction, validation, and simulation of hits in the CMS Pixel Detector. Proceedings of Science: Vertex.2007:35. July 2007

Work Experience

Research Assistant - JHU - 2014-present

Teaching Assistant - JHU - 2014-2015

Representative, Inbound Sales and Support - GoDaddy - 2012-2014